

Intrinsic Disorder and Allostery in Glucocorticoid Receptor

by

Jing Li

A dissertation submitted to Johns Hopkins University in conformity with
the requirements for the degree of Doctor of Philosophy

Baltimore, Maryland

May 2014

Abstract

Intrinsically disordered (ID) regions of proteins, lacking stable tertiary structure, are malleable and sensitive regulators of cell functions. Allostery is transmittance of a perturbation at one region to distant sites of the same molecule, allowing for precise control of macromolecular function. Nature preferentially uses both ID regions and allostery to regulate protein function, as observed in transcription factors. This motivated us to investigate whether and how ID regions can facilitate allostery.

The classic allosteric models all feature a static structural and single molecule view of allostery and are not applicable to study allostery mediated by ID regions. To address this limitation, our group has developed the Ensemble Allosteric Model (EAM), which views allostery as an effector binding driven shift in ensemble probabilities. It is a thermodynamic and quantitative model that can be applied to structured proteins, ID proteins and mixed proteins.

In this thesis project, I investigated the intra- and inter- domain allostery mediated by the ID regions in human glucocorticoid receptor (GR). Through thermodynamic and functional studies on eight GR translational isoforms, I found that the ID GR N terminal domain (NTD) is composed of two functionally distinct regions, unfavorably coupled with each other, and both of them are favorably coupled to the DNA binding domain (DBD).

Based on these experimental constraints, an EAM was built for GR and reasonable thermodynamic parameter combinations were found to describe both the transcriptional activity and binding affinity of different translational isoforms. We found that GR uses these competing energetic couplings, which are modulated in different translational isoforms, to provide tunable responses to environmental cues.

In the context of the EAM predictions, mutagenesis was carried out in different regions of GR and the influence on transcriptional activity and binding affinity was assayed to pinpoint the molecular basis of allostery, which paves a way for allosteric drug design. This study suggests a unifying strategy to investigate thermodynamics and the molecular basis of allostery in any complicated system.

Preface

There are many people who provided me mentorship, assistance, encouragement and love through my entire graduate school. I would like to take the opportunity here in the preface to thank all of them.

First of all, I would like to thank my thesis committee members in Jenkins: Dr. Greg Bowman, Dr. Richard Cone, Dr. George Rose and Dr. Doug Barrick. In each thesis review they gave me a lot of helpful suggestions and guidance. I gratefully thank them for their mentorship.

My mentor Dr. Vincent Hilser deserves my eternal respect and thanks. He continuously gives me intellectual support, as well as freedom to pursue my project during these five years. He is such an encouraging and considerate mentor who helped me through each challenge I encountered. The most important thing I learned from him is to how to be a confident and enthusiastic scientist, and how to write and present science. The other scientist in our lab who deserves my heartfelt thanks is Dr. James Wrabl. He is the one I can consult anytime about any technical questions or experimental designs. He is always there for any student in the lab. I would also like to thank all previous and present Hilser lab members particularly Dr. Travis Schrank, Dr. William Elam, Harry Saavedra, Jordan White, Hesam Motlagh, Andrew Martens, Jeremy Anderson and Alex Chin for their feedback on my project. In addition, I deeply appreciate the undergraduates James Sowers, Kaixian Liu and Carolyn Chakuroff's participation in my project.

My dissertation would not be possible without the support and help of our collaborators, Dr. E. Brad Thompson, Dr. Trina Schroer, Dr. Doug Barrick and Dr. Bertrand Garcia-Moreno. I deeply appreciate their sharing instruments with us, and providing advice and feedback on my project.

At last, I would like to acknowledge the people in my personal life, my parents, Weimin Li and Guilan Sun, my husband, Jiabin Yan, and my daughter, Sophia Yan. Without their love, support and encouragement, I would not be able to pursue my graduate study here.

Table of Contents

Abstract.....	ii
Preface.....	iv
Table of Contents.....	vi
List of Abbreviations.....	xiii
Chapter 1- Introduction.....	1
1.1 Allostery, classic allosteric models and the Ensemble Allosteric Model (EAM)	1
1.2 The importance of intrinsically disordered proteins	2
1.3 The interplay between allostery and intrinsic disorder manifested in EAM	2
1.4 Glucocorticoid receptor, an experimental system to investigate the interplay between intrinsic disorder and allostery from the ensemble view	3
1.5 Overview of thesis.....	4
Figure 1 Domain architecture (top) and organization (bottom) of the eight, two-domain glucocorticoid receptor isoforms.....	7
Chapter 2- Thermodynamic Dissection of the Intrinsically Disordered N-terminal Domain of Human Glucocorticoid Receptor	8
2.1 Abstract	8
2.2 Introduction	8
2.3 Materials and methods.....	10
<i>Protein expression, purification and storage</i>	10
<i>Preparation of Trimethylamine N-oxide (TMAO) buffers</i>	12
<i>TMAO induced protein folding transition monitored by tryptophan emission fluorescence intensity</i>	12
<i>Protease protection assay</i>	13
2.4 Results and discussion	13
<i>Expression and purification of GR NTDs in E. coli</i>	13
<i>The naturally occurring osmolyte TMAO can induce cooperative folding transitions in the NTDs of different GR translational isoforms</i>	14
<i>Protease protection assay validates the rank order of stability estimates from TMAO folding of the GR A-NTD, GR C2-NTD, and GR C3-NTD constructs</i>	14
<i>The folded states of GR NTDs induced by TMAO are thermodynamically similar to globular proteins of similar length</i>	15
<i>Full length GR NTD contains at least two thermodynamically coupled regions</i>	16
2.5 Conclusions	18
Figure 2 Schematic representation of the GR's NTD sequence showing the positions of different GR translational isoform start sites.	19
Figure 3 TMAO-induced conformational transitions of GR A-NTD, C2-NTD, and C3-NTD monitored by tryptophan emission fluorescence intensity.	20
Figure 4 Protease protection assays for GR A-NTD, C2-NTD, and C3-NTD.	21
Figure 5 Dependence of m-values on protein chain length (L) for TMAO induced folding transitions.	22
Figure 6 Schematic representation of the ensemble allosteric model of full-length GR NTD containing the R domain and F domain.....	23

Figure 7 Circular dichroism data comparison for globular and ID protein in terms of $\Delta\epsilon_{222\text{ nm}}$ versus $\Delta\epsilon_{220\text{ nm}}$	24
Figure 8 Dependence of the population of the states with F domain folded on TMAO concentration for A-NTD and C3-NTD.....	25
Table1. Fitting parameters for GR A-NTD, GR C2-NTD and GR C3-NTD constructs from TMAO induced folding transitions.....	26
Chapter 3- Thermodynamic Dissection of the NTD and DBD Two-domain Constructs of Human Glucocorticoid Receptor	27
3.1 Abstract	27
3.2 Introduction	27
3.3 Materials and methods.....	29
<i>Plasmids</i>	29
<i>Protein expression, purification and storage</i>	29
<i>Osmolyte TMAO induced protein folding transitions</i>	29
<i>Protease protection assay</i>	29
3.4 Results and discussion	30
<i>TMAO induced folding of GR A-NTDDBD construct does not follow cooperative two-state transition</i>	30
<i>TMAO could induce cooperative folding transitions in C3-NTDDBD construct and the folded conformation of its ID part is stabilized in presence of DBD</i>	31
<i>Protease Protection Assay suggests allosteric coupling between R domain and DBD</i>	32
3.5 Conclusions	33
Figure 9 Fluorescence emission spectrum of GR A-NTDDBD at 0M TMAO vs. 2.5 M TMAO.	34
Figure 10 Comparison of TMAO-induced conformational transitions of GR A-NTDDBD and GR A-NTD monitored by tryptophan emission fluorescence intensity.	35
Figure 11 TMAO-induced conformational transitions of GR C3-NTDDBD and GR C3-NTD monitored by tryptophan emission fluorescence intensity.	36
Figure 12 Mapping m-value for GR A-NTDDBD and GR C3-NTDDBD on the m-value versus protein chain length chart.	37
Figure 13 Protease protection assays for GR A-NTDDBD and GR C3-NTDDBD.	38
Table2. Fitted parameters for GR A-NTD, GR A-NTDDBD, GR C3-NTD, and GR C3-NTDDBD constructs from TMAO induced folding transitions.	39
Chapter 4- Binding Energetics of the Two-domain Constructs of GR Translational isoforms to GR Response Element	40
4.1 Abstract	40
4.2 Introduction	40
4.3 Materials and methods.....	41
<i>Protein expression, purification and storage</i>	41
<i>6FAM labeled half site GRE and control oligonucleotide</i>	41
<i>Binding affinity measurement monitored by fluorescence anisotropy change</i>	41
<i>Binding affinity fitted with single site binding model</i>	42
4.4 Results and discussion	42

<i>Eight GR translational isoforms have variable binding affinities to GRE</i>	42
4.5 Conclusions	43
Figure 14 Pattern of Glucocorticoid response element (GRE).....	44
Figure 15 Illustration of experimental design for binding GR's two-domain constructs to 6FAM labeled GRE.	45
Figure 16 Binding of NTDDDBD two-domain constructs of eight GR translational isoforms to GRE monitored by fluorescence anisotropy change.....	46
Figure 17 Bar plot comparison of eight GR translational isoforms' binding affinity to half site GRE.	47
Table3. Fitted parameters for binding between NTDDDBD two-domain constructs of eight GR translational isoforms to 6FAM labeled GRE.....	48
Chapter 5-Cell Based Functional Assay to Measure the Transcriptional Activity and Relative Binding Affinity of GR.....	49
5.1 Abstract	49
5.2 Introduction	49
5.3 Materials and methods.....	50
<i>Mammalian cell expression vectors for the NTDDDBD constructs of GR translational isoforms</i>	50
<i>Luciferase reporter vectors</i>	51
<i>Mammalian cell culture and transfection</i>	51
<i>Western blots</i>	51
<i>Immunostaining</i>	53
<i>Luciferase dosage curve to measure transcriptional activity of GR translational isoforms</i>	53
<i>Competitive transfection assay to measure the relative binding affinity</i>	54
5.4 Results and discussion	55
<i>U-2 OS cells transiently transfected with each isoform can express each isoform by itself</i>	55
<i>Different GR translational isoforms have similar nuclear translocation efficiencies</i>	56
<i>Correlation between the binding affinity measured in vitro and the EC50 fitted from luciferase dosage curve</i>	56
<i>Competitive transfection assays can approximate the binding affinity of the inactive isoforms</i>	57
<i>Lack of correlation between the transcriptional activity and binding affinity of eight GR translational isoforms</i>	58
5.5 Conclusions	59
Figure 18 Co-transfection strategy in the dual luciferase reporter assay.....	60
Figure 19 U-2 OS cells transiently transfected with each isoform can express each single isoform shown by Western blot.	61
Figure 20 Immunostaining of U-2 OS cells transfected with GR translational isoforms.....	62
Figure 21 Dosage curve of the transcriptional activity of the two-domain constitutively active constructs of the eight translational isoforms monitored by dual luciferase reporter assay.....	63
Figure 22 Correlation between EC50 fitted from <i>in vivo</i> dosage curve and <i>in vitro</i>	

measured binding affinity for five active GR translational isoforms.	64
Figure 23 Competitive binding luciferase assay comparing increasing amounts of D1, D2, D3 and DBD co-transfected with the C3 isoform	65
Figure 24 Correlation between the EC50 fitted from the competitive binding assay and the <i>in vitro</i> measured binding affinity for D1, D2 and D3 isoforms.	66
Figure 25 Lack of correlation between transcriptional activity and binding affinity to GRE of the eight GR translational isoforms.	67
Table 4. Fitted parameters from the luciferase assay dosage curve for A, B, C1, C2 and C3 five active isoforms.	68
Chapter 6-Bidirectional Competing Allosteric Coupling in GR	69
6.1 Abstract	69
6.2 Introduction	69
6.3 Materials and methods.....	70
<i>Constructs</i>	70
<i>Competitive transfection assay</i>	70
6.4 Results and discussion	70
<i>Directly linking R domains of A and B isoforms to DBD significantly increases DBD's binding affinity to GRE</i>	70
<i>Bidirectional energetic couplings are competitive in constitutively-active, two-domain GR constructs</i>	71
6.5 Conclusions	72
Figure 26 Competitive transfection assay comparing a construct with the R domain conjoined to the DBD versus the DBD construct.....	73
Figure 27 Summary of the allosteric coupling network in the two-domain construct of GR.	74
Figure 28 Bi-directional competing allosteric coupling network when allosteric effector GRE binds to the DNA binding domain.....	75
Chapter 7-Simulation with Ensemble Allosteric Model for GR Translational Isoforms	76
7.1 Abstract	76
7.2 Introduction	76
7.3 Methods	77
<i>EAM for each two-domain construct of GR translational isoforms</i>	77
<i>Parameter space search with EAM to satisfy experimental constraints</i>	78
7.4 Results and discussion	78
<i>The competing energetic coupling is crucial to satisfy all experimental constraints</i> .	78
<i>EAM can capture the absence of a correlation between the transcriptional activity and binding affinity for GR translational isoforms</i>	78
7.5 Conclusions	79
Figure 29 Ensemble allosteric model for A and C3 isoforms.....	80
Table 5. Practical parameter range used in parameter space search	81
Table 6. Experimental constraints for relative activity and binding affinity for each isoform with full length A isoform as reference in the parameter space search	82
Figure 30 Box chart of satisfied parameter range given experimental constraints on the Ensemble Allosteric Model.	83

Figure 31 Correlation between the simulated relative transcriptional activity and binding affinity with experimentally measured values.	84
Simulated values for transcriptional activity and binding affinity of each isoform are based on the ensemble averaged contribution of all the states in the ensemble. The correlation coefficient between the simulated data and experimental data is the Pearson correlation.	
84	
Chapter 8- Applications of EAM: Predictions of Mutational Effects.....	85
8.1 Abstract	85
8.2 Introduction	85
8.3 Methods	86
<i>Perturbation of GR EAM with mutagenesis introduced to R domain, F domain or DBD</i>	86
8.4 Results and discussion	87
<i>Perturbations of GR EAM with mutagenesis introduced to R domain of different translational isoforms</i>	87
<i>Perturbations of GR EAM with mutagenesis introduced to F domain of different translational isoforms</i>	88
<i>Perturbations of GR EAM with mutagenesis introduced to DBD of different translational isoforms</i>	89
8.5 Conclusions	90
Figure 32 Predicted influence on transcriptional activity and binding affinity to GRE when mutations are introduced to R domain in different isoforms.....	91
Figure 33 Predicted influence on transcriptional activity and binding affinity to GRE when mutations are introduced to F domain in different isoforms.....	92
Figure 34 Predicted influence on transcriptional activity and binding affinity to GRE when mutations are introduced to DBD in different isoforms.	93
Chapter 9-Molecular Basis for the Unfavorable Coupling between the Regulatory (R) and Functional (F) Domains in GR NTD	94
9.1 Abstract	94
9.2 Introduction	94
9.3 Materials and methods.....	95
<i>Site directed mutagenesis</i>	95
<i>Western blot</i>	95
<i>Transcriptional activity measurement at the saturated GR vector level</i>	96
9.4 Results and discussion	96
<i>Mutations within segment 90-97 dramatically increase transcriptional activity on C2 isoform</i>	96
<i>Mutations on the F domain have different functional effects on A and C2 isoforms compared with C3 isoform</i>	98
9.5 Conclusions	99
Figure 35 Most point mutations within segment 90-97 significantly increase the transcriptional activity compared to wild type C2.....	101
Figure 36 Mutations within R domain on A, B, C1 and C2 isoforms to identify residues involved in the R and F coupling network.	102

Figure37 Flexible linkers with varied length flanking the N termini of C3 significantly decrease the transcriptional activity.....	103
Figure 38 Mutations within F domain on A, C2 and C3 isoforms to identify residues involved in the R and F coupling network.	104
Chapter 10-Molecular Basis of the Favorable Coupling between the Functional (F) domain in the NTD and DBD	105
10.1 Abstract	105
10.2 Introduction	105
10.3 Materials and methods.....	106
<i>Constructs</i>	106
<i>Western blot</i>	106
<i>Transcriptional activity measurement</i>	106
10.4 Results and discussion	106
<i>Chimeric construct linking F domain to yeast transcription factor Gal4's DBD showed significantly decreased transcriptional activity compared to natural GR construct.</i>	106
<i>Mutagenesis on F domain reveals some residues involved in the coupling network</i>	107
<i>Mutagenesis on DBD pinpoints some residues involved in coupling to F domain...</i>	108
10.5 Conclusions	109
Figure 39 Domain swapping of GR DBD to Gal4 DBD significantly decrease transcriptional activity.....	110
Figure 40 Transcriptional activity influence of single point mutations within F domain tested on C3 isoform.....	111
Figure 41 Mutations within F domain on C3 isoform and chimeric construct (F domain tethered to Gal4 DBD) to identify residues involved in the F and DBD coupling network	112
Figure 42 Sequence conservation analysis of DBD within GRs versus SHRs.....	113
Figure 43 Influence on transcriptional activity of single point mutations carried out within DBD on C3 isoform.	114
Figure 44 Mutations within DBD on C3 isoform to identify residues involved in the F and DBD coupling network.	115
Chapter 11- Molecular Basis for the Favorable Coupling between the Regulatory (R) Domain in the NTD and DBD.....	116
11.1 Abstract	116
11.2 Introduction	116
11.3 Materials and methods.....	117
<i>Constructs</i>	117
<i>Competitive transfection assay</i>	117
11.4 Results and discussion	117
<i>Segment within R domain potentially involved in the R domain and DBD coupling network</i>	117
<i>Residues within DBD potentially involved in the R domain and DBD coupling network</i>	118
11.5 Conclusions	119
Figure 45 Truncations of the construct with the R domain joined to the DBD to identify	

segment involved in the R domain and DBD coupling network.	120
Figure 46 Mutations within DBD and R _A -DBD construct to identify residues involved in the R and DBD coupling network.....	121
Figure 47 EC50 of DBD and R _A -DBD constructs' mutants fitted from competitive transfection assay.	122
Chapter 12- Concluding Remarks	123
12.1 Summary	123
12.2 Toward future applications.....	123
Figure 48 Cartoon summary of the molecular basis of allosteric coupling in GR constitutively active two-domain constructs	125
Reference.....	126
Appendix 1. Derivation of the fitting equation for GRE and GR binding	130
Appendix 2. C code for GR EAM model.....	131
Bibliography	141

List of Abbreviations

A:	alanine
AF1:	activation function 1
C:	cysteine
D:	aspartic acid
DBD:	DNA binding domain
E:	glutamic acid
EAM:	ensemble allosteric model
6-FAM:	6-Carboxyfluorescein
F domain:	functional region
G:	glycine
GR:	glucocorticoid receptor
GRE:	glucocorticoid response element
HRE:	hormone response element
ID:	Intrinsically disordered
IDP:	intrinsically disordered protein
IDR:	intrinsically disordered region
IPTG:	isopropyl-D-thiogalactopyranoside
K:	lysine
KNF:	Koshland-Nemethy-Filmer
L:	leucine
LBD:	ligand binding domain

LEM:	linear extrapolation method
MWC:	Monod-Wyman-Changeux
N:	asparagine
NTD:	N terminal domain
PBS:	Phosphate buffered saline
Q:	glutamine
R:	arginine
SDS-PAGE:	sodium dodecyl sulfate polyacrylamide gel electrophoresis
R domain:	regulatory domain
SHR:	steroid hormone receptor
S.W.:	statistical weight
T:	threonine
TMAO:	trimethylamine-N-oxide
V:	valine

Chapter 1- Introduction

1.1 Allostery, classic allosteric models and the Ensemble Allosteric Model (EAM)

Allostery is the process by which remote sites of a system are energetically coupled to elicit a functional response, which manifests itself as “action at a distance”. It plays important roles in signal transduction, gene transcription and metabolic homeostasis by endowing the cells with a sensitive responsiveness to environmental cues [1]. Due to its functional importance, understanding the structural, molecular and energetic basis of allostery in individual systems is crucial to dissecting the mechanism. Historically, the structural view of allostery has guided the field, represented by two classic models, the concerted MWC (Monod-Wyman-Changeux) model[2] and the sequential KNF (Koshland-Nemethy-Filmer) model [3]. The characteristic feature of the MWC model is that the conformational change between the two domains is coupled. In contrast, the classic feature of the KNF model is that conformational change and ligand binding are coupled [4]. Both of these two models assume a static structural view of allostery. However, accumulating evidence has called into question these classic views of allostery, especially the hyper-abundance of intrinsic disorder in allosteric regulatory proteins, such as transcription factors [5]. Recently, the Ensemble Allosteric Model (EAM) [6-8], which is pioneered by our lab, may provide insights into these systems. In this model, the molecule without effectors is considered as an ensemble of pre-existing conformations, and effector binding is treated as an energetic perturbation of the ensemble to redistribute the microstates that are favorable or unfavorable to the second binding partner. It only considers the intrinsic energetics of the system, does not depend on a crystallographic structure, and is more general than the classic models. All the allosteric coupling discussed in this dissertation is based on the ensemble allosteric concept.

1.2 The importance of intrinsically disordered proteins

Intrinsically disordered proteins (IDPs) lack a stable tertiary structure under physiological conditions and exist as ensembles of different conformational states. More than 30% of the proteome is thought to consist of IDPs [9] and intrinsically disordered regions (IDRs) are hyper-abundant in transcription factors and cell signaling proteins [5]. IDPs have challenged the dominant view of protein structure-function relationships. The following are some of the unique characteristics of intrinsically disordered regions (IDRs): high specificity for multiple targets due to structural plasticity, different conformations in the bound state depending on the binding partners, high specificity and relatively low binding affinity to binding partners, fast association and dissociation rates with binding partners [10], and enrichment with post-translational modification sites [11] and splicing isoforms [12]. All these attributes together make IDRs tunable and versatile for regulation of cellular functions.

1.3 The interplay between allostery and intrinsic disorder manifested in EAM

The coexistence of allostery and intrinsic disorder in regulatory proteins, such as cell signaling proteins and transcription factors, suggests that IDRs may play important roles in regulating allosteric coupling. Our group first entered this field by proposing a simple ensemble model [6] to study allosteric coupling. The model treated two or three tethered domains in an allosteric protein as being either folded (high affinity conformation to ligand) or unfolded (low affinity conformation to ligand), resulting in multiple states in the ensemble. Coupling between domains is expressed as interaction energy, which represents the statistical nature of the interaction. Depending on the energetic costs of interconverting between the folded state and unfolded state of each domain and of breaking the interaction between domains, each state in the

ensemble has a different probability. Effector binding stabilizes the high affinity state to effector and thus redistributes the ensemble. Allosteric coupling response is considered as the degree to which the probability of states in which functional sites are folded is affected for a given perturbation to states that can bind effectors. An unbiased and exhaustive search of parameter space found that allosteric coupling is maximized when the proteins contain ID domains and significant interaction energy between the two domains is a prerequisite for coupling [6]. More generally, allostery caused by different energetic perturbations (besides ligand binding) to the preexisting ensemble, such as post-translational modifications, mutations, and truncations can all be understood in terms of the EAM [6].

1.4 Glucocorticoid receptor, an experimental system to investigate the interplay between intrinsic disorder and allostery from the ensemble view

Glucocorticoid receptor (GR), a transcription factor in the steroid hormone receptor (SHR) family, plays key roles in organ development, metabolite homeostasis and the inflammatory response [13]. Malfunction of GR leads to cancer, metabolic diseases and inflammatory disorders [14], and GR-based pharmaceuticals are among the most commonly prescribed drugs. GR contains three major domains, the intrinsically disordered N-terminal domain (ID NTD), the DNA binding domain (DBD) and the ligand binding domain (LBD) [15]. Allostery is known to govern the function of SHRs [16, 17]. In GR, the intra-domain and inter-domain allosteric couplings are widely reported in the literature [18-22]. However, the mechanism is elusive, especially how the long NTD is involved in the coupling due to the difficulty of investigating it. The major transactivation region of GR, activation function 1 (AF1) region, which contains co-regulators' binding sites, is located in the NTD [23, 24]. In addition the functional importance of GR's NTD

is also suggested by the recently discovered eight human GR translational isoforms with different tissue distributions, different transcriptional activities and unique gene regulation sets [25]. Surprisingly, these translational isoforms differ only in the length of their NTD [25, 26] (as shown in Fig. 1). All of these led us to investigate whether and how the NTD is involved in the allosteric coupling network in GR. Truncation of the LBD removes GR's transcriptional activity dependence on hormone and generates a constitutively active, two-domain construct [27] as shown in Fig. 1. The constitutively active constructs of different translational isoforms were used for functional assays in this thesis.

1.5 Overview of thesis

In Chapter 2, the osmolyte TMAO was used to induce folding transitions of the NTD of three representative translational isoforms, A, C2 and C3. It was found that the full length NTD can be viewed as having at least two thermodynamically coupled regions, a functional (F) domain, indispensable for GR transcriptional activity, and a regulatory (R) domain, whose length serves to regulate the activity of GR [26] (as shown in Fig. 1a). These two regions are unfavorably coupled to each other.

Chapter 3 compares osmolyte induced folding transitions of the single NTD construct and the NTD-DBD two-domain constructs of the A and C3 isoforms. We found that DBD influences both the folding cooperativity and the energetics of the NTD. This suggested that the NTD and DBD were allosterically coupled to each other.

In Chapter 4, binding affinities to GRE were measured for the eight translational isoforms. They show different binding affinities varying about 5 fold even though they have exactly the same DBD. This suggests that the ID NTD is allosterically coupled to DBD, and in different

translational isoforms, the coupling is differentially regulated. The binding affinities of the eight isoforms provide useful hints about the allosteric coupling mechanism between the ID NTD and DBD, and guided us to generate testable hypothesis in the functional assays.

For Chapter 5, a dual luciferase reporter assay in a mammalian cell line, U-2 OS, was used to test the hypothesis generated from the biophysical studies. Dependence of transcription activity on GR vector dosage was measured for each translational isoform, and maximum transcriptional activity and relative binding strength to GRE of each isoform was fitted. In addition, a competitive transfection assay was designed and shown to be valuable for measuring the binding of inactive GR isoforms to the GRE. Transcriptional activity was tunable for the eight translational isoforms and did not correlate with the binding affinity to GRE. This observation was puzzling at first, because the binding of transcription factors to their response elements generally determines their specificity and activity [28, 29]. In the remaining part of this thesis, the detailed allosteric coupling mechanism was investigated to resolve this puzzling observation.

In Chapter 6, the hypothesis that the R domain is favorably coupled to the DBD was further tested using the luciferase assay. Different parts of the R domain were directly linked to DBD through a flexible linker and the influence on DBD's binding affinity to GRE was measured through the competitive transfection assay. The results supported the favorable coupling between the R domain and DBD. The thermodynamics of allostery in GR can be summarized based on the *in vitro* biophysical characterization and the cell based functional assay. Bidirectional competing energetic coupling is found to tune the transcriptional activities and binding affinities of different translational isoforms.

In Chapter 7, the ensemble allosteric model (EAM) was built for eight GR translational

isoforms based on the experimental constraints. Reasonable thermodynamic parameter combinations could be found such that the simulated transcription activity and binding affinity are both significantly correlated with the experimental measurements. The lack of correlation between transcriptional activity and binding affinity of eight GR translational isoforms could also be captured by the EAM.

Mutagenesis is commonly used to probe which residues are involved in an allosteric coupling network. However, in complicated systems, the mutational effects are difficult to interpret. Based on the EAM model we constructed for GR, the influence on transcriptional activity and binding affinity upon mutagenesis can be predicted for different scenarios. This is discussed in Chapter 8, and paves a way for pinpointing the molecular basis of allostery in GR.

Chapter 9 describes how the molecular basis of allostery between the R and F domains was investigated by site directed mutagenesis in luciferase activity assays, and some residues involved in the coupling network were found.

In Chapter 10, we explored the coupling mechanism between the F domain and DBD. Point mutations and chimeric constructs of the C3 isoform were tested using the luciferase activity assay. All the results support that there is favorable coupling between the F domain and DBD, and the molecular basis for the coupling was interpreted from these experiments.

In Chapter 11, to investigate the molecular basis of allostery between the R domain and DBD, single point mutations were introduced to the DBD construct and a construct with the R domain conjoined with the DBD. Interpreting mutational effects based on EAM predictions helped us to pinpoint the coupling mechanism.

In Chapter 12, the importance and possible application of the research is discussed.

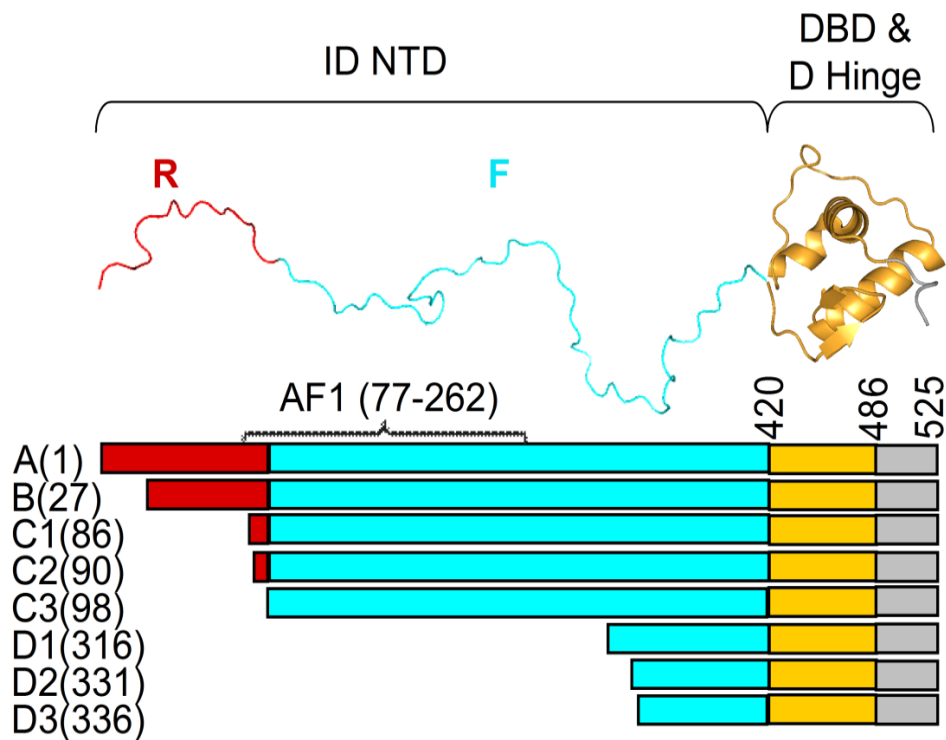


Figure 1 | Domain architecture (top) and organization (bottom) of the eight, two-domain glucocorticoid receptor isoforms.

Human glucocorticoid receptor (GR) constructs that are constitutively active are composed of the intrinsically disordered N-terminal domain (NTD) (1-420), and the DNA binding domain (DBD) (421-486), as well as part of hinge region (487-525), is shown as cartoon. The NTD is subdivided into the regulatory (R) domain and functional (F) domain, There are 8 translational isoforms, A, B, C1, C2, C3, D1, D2 and D3 with differing lengths of their NTD. The lower part of the graph systematically shows the starting sites of the isoforms labeled according to their position in the full-length receptor. Activation function 1 (AF1) region (77-262) is also mapped on the figure.

Chapter 2- Thermodynamic Dissection of the Intrinsically Disordered N-terminal Domain of Human Glucocorticoid Receptor

2.1 Abstract

Intrinsically disordered (ID) regions are abundant in cell signaling proteins and transcription factors. As ID regions commonly fold up during intracellular function, it is crucial to understand these folded states and how they affect the overall function of the parent protein. In this part, we thermodynamically characterized the intrinsically disordered ID N-terminal domains (NTD) of three glucocorticoid receptor (GR) translational isoforms by using the osmolyte trimethylamine N-oxide (TMAO) to induce folding transitions. All the three investigated NTD isoforms were found to be cooperatively foldable to globular protein-like conformers in TMAO. In addition, the full length NTD was found to contain at least two functionally distinct regions, the regulatory (R) domain and the functional (F) domain, which are negatively coupled to each other. Remarkably, adjusting the length of R (as observed in naturally occurring isoforms), turned out to regulate the stability of the NTD. This suggests that regulating the stability of ID regions could be used as a paradigm for regulating the function of ID proteins.

2.2 Introduction

The classical view of protein structure-function relationships, the protein structure defines its function, has been challenged by the increasing number of proteins found to contain ID regions, particularly cell signaling proteins and transcription factors [5, 30, 31]. Many of these ID regions are found to contain abundant binding sites for important protein partners [30, 32-34], and folding coupled with binding usually happens when these ID regions encounter their binding partners [30, 32-34]. Thus in most cases, the conditionally folded states are the functional states of ID proteins.

Mutagenesis [35], truncation or translocation of ID regions in some proteins [32] is usually involved in diseases. However, largely unknown are the mechanisms by which these changes in ID regions regulate protein functions. Osmolyte induced folding of some ID proteins has been reported and the induced folded states were determined to be functionally relevant [36, 37]. This provides a way to study the folded state of ID proteins. However, the folding cooperativity, judged by m-value in osmolyte induced folding [38], is poorly characterized. Amino acid usage in ID proteins is significantly different from that in globular proteins [31]. Whether the thermodynamic ground rules governing globular proteins can be used in studying ID proteins is not clear.

To better understand proteins containing ID regions, we carried out thermodynamic characterization of the ID N-terminal domains (NTD) of three human GR translational isoforms. The NTDs for SHRs are extremely important for transcriptional regulation, serving as a hub to recruit co-regulators to form the final transcriptional complex [13, 24, 39]. The “hormone-independent” transactivation domains AF1 have been mapped within each SHR NTD by deletion, insertional mutation or point mutagenesis analysis [24]. In human GR, AF1 is mapped onto residues 77-262 [40] (as shown in Fig. 2). Recently, different human GR N-terminal translational isoforms with differing activities, tissue distributions and unique gene regulation sets were reported to be generated from one single GR mRNA through translational regulatory mechanisms [25]. These translational isoforms only differ in the length of their NTDs (as shown in Fig. 1 and Fig. 2). Except for the isoforms (D1, D2 and D3) that do not overlap any part of the AF1 transactivation domain, these isoforms are all active and have different potencies in transcriptional regulation [25]. Especially intriguing was how the activity is highly sensitive with as little as eight amino acid differences in the ID domain (i.e. the length difference between C2

and C3 isoforms) playing a key role in functional regulation. We carried out the study of human GR using the A (amino acid 1-420), C2 (90-420), and C3 (98-420) isoforms.

To access the folded conformations of these ID proteins, the naturally occurring osmolyte trimethylamine N-oxide (TMAO) was employed to induce the folding transitions. Organic osmolytes are small molecules in cells and function to stabilize and protect intracellular proteins against common denaturing environmental stresses [41, 42]. TMAO has been demonstrated to force thermodynamically unstable proteins to fold and regain high functional activities [18, 36, 43].

In this chapter, we report a novel strategy to express and purify the long ID GR NTD in *E. coli*, enabling measurement of the cooperative folding transitions of GR NTDs to globular protein-like folded conformations when induced by TMAO. The results demonstrate that GR NTD contains at least two thermodynamically distinct regions, the “regulatory (R) domain” and the “functional (F) domain”, which are unfavorably coupled to each other.

2.3 Materials and methods

Protein expression, purification and storage

Coding sequences to express GR A-NTD, GR C2-NTD and GR C3-NTD with 10 tandem histidines tagged on both N and C termini were optimized for *E. coli* expression and synthesized by DNA 2.0 (Menlo Park, CA), and inserted into the pJ411 bacterial expression vector under T7 promoter control. BL21(DE3)pLysS competent cells (Novagen) transformed with expression plasmids were plated on LB plates with 30ug/mL kanamycin and incubated at 37°C overnight. A single colony was picked up and inoculated into 50 mL LB medium with 30ug/mL kanamycin and grown in 30°C incubator shaker at 250 rpm overnight. The following morning, each 10mL

overnight culture was transferred to 500mL LB medium and grown at 37°C to OD600 0.8, then the temperature was adjusted to 15°C. After 1 hour, the culture was induced by 1 mM isopropyl- β -D-thiogalactopyranoside (IPTG). After 18 hours at 15°C with shaking at 250rpm, *E. coli* cells were collected by centrifugation for 15min at 5000 rpm at 4 °C and pellets were washed with phosphate buffered saline (PBS) and centrifuged again. To cell pellets from 2 L culture, 50 mL denaturing lysis buffer (6M guanidine hydrochloride, 100mM monosodium phosphate, 10mM Tris, 20mM imidazole, pH 8.0) and one cOmplete- EDTA-free ® protease inhibitor cocktail tablet (Roche) were added. Cells were lysed by flash freezing for 20 min twice in a dry ice and ethanol mixture and thawing for 10 min in 42°C water bath twice. Cell lysate was centrifuged at 30,000g at 4 °C for 1 hour. The supernatant from 4 L induced cells was loaded onto LPLC column packed with 20 mL Ni-NTA superflow resin (Qiagen) pre-equilibrated with the denaturing lysis buffer by gravity flow. After thoroughly washing the column with the denaturing lysis buffer at 2 mL/min, protein was renatured on the column by flowing native lysis buffer (100mM monosodium phosphate, 10mM Tris, 20mM imidazole, pH 8.0). Afterwards, native lysis buffer containing 200mM imidazole was applied to the column at 2 mL/min, to wash out the contaminants and degradation products with His-tag only on one terminus. Finally, target protein was eluted from the column by native lysis buffer containing 500mM imidazole. Protein purity was checked on SDS-PAGE. The purified target protein was dialyzed against storage buffer containing 10mM HEPES, 80mM NaCl, 1mM EDTA and 10% glycerol, pH 7.6. Protein concentration was then determined by the absorbance at 280 nm in the UV spectrometer according to the Edelhoch method [44]. Each 500uL protein aliquot in 1.5 mL Protein Lobind Tube (Eppendorf) was then flash frozen in an ethanol- dry ice bath and stored at -80°C.

Preparation of Trimethylamine N-oxide (TMAO) buffers

Trimethylamine-N-oxide Dihydrate (TMAO, 98% pure, Acros Organics) was dissolved in 100mM Tris, 200mM NaCl, 50mM Arginine buffer to make the 0M, 0.5M, 1M, 1.5M, 2M and 2.5M TMAO buffers. pH was adjusted to 7.4 for each buffer separately. Activated carbon, 12-20 mesh (Sigma-Aldrich) was added into each buffer solution and stirred for 30min while protected from light to absorb the impurities. The buffer was then filtered (0.22um filter, Millipore), aliquoted, and stored at -80°C. TMAO buffers with similar concentrations were mixed to obtain the target TMAO concentration used in the following fluorescence measurements, minimizing the pH changes due to mixing two TMAO buffers with equal pH but large concentration difference. Arginine, which is reported to aid protein solubility [45], was included in the TMAO buffers at a concentration of 50mM and efficiently prevented protein aggregation.

TMAO induced protein folding transition monitored by tryptophan emission fluorescence intensity

Fluorescence emission spectra for each purified recombinant GR NTD construct were measured using an Aviv ATF 105 fluorometer (Aviv Biomedical) in TMAO buffers of varying concentrations. Freshly prepared dithiothreitol 1 M stock solution was added to TMAO buffer to a final concentration of 1 mM. Protein at 0.5 μ M in a volume of 150 μ L TMAO buffer at 22°C was allowed to rest in a “sub micro” fluorometer cell (Santa Cells) for 5 minutes to allow for temperature stabilization and protein conformational equilibrium to be reached. Emission spectra were then recorded with excitation at 295nm. All spectra were corrected for the contribution of its corresponding buffer.

Fluorescence emission intensities at 338nm were recorded as a function of TMAO

concentration and normalized to the intensity at 0M TMAO concentration. The resulting sigmoid curve was fitted to a two-state cooperative folding transition with linear extrapolation method (LEM, [46]) to determine the stability ($\Delta G_{U \rightarrow F}$) and m-value for each construct. The LEM method has been shown to be valid for fitting the data of osmolyte induced folding of ID proteins [47].

Protease protection assay

Protease digestions of purified GR A-NTD, GR C2-NTD and GR C3-NTD proteins at 1 mg/mL were performed with sequencing grade trypsin (Sigma) at 22°C. Digestions were performed at a protein versus trypsin mass ratio of 1000:1 in 10 mM HEPES, 80 mM NaCl, 1 mM EDTA and 10% glycerol buffer, pH 7.6 for 0 min, 5 min, 10 min and 30 min. Digestions were quenched by mixing protein and trypsin mixtures with 6x Laemmli sample buffer and boiling at 100°C for 10min. 40 µg of sample at each time point was separated on 4-15%, SDS-PAGE gels (Bio-Rad) with SDS-Tris-Glycine gel running buffer.

2.4 Results and discussion

Expression and purification of GR NTDs in E. coli

Expression and purification of GR A-NTD, GR C2-NTD and GR C3-NTD constructs were carried out in *E. coli*. As degradation and low yield are general problems in expression and purification of recombinant ID proteins from *E. coli*, I used the following: engineered plasmids with optimized codons for *E. coli* (DNA2.0), affinity purification tags on both the N terminus and C terminus of each construct, a protease deficient competent expression strain (BL21(DE3)pLysS), overnight protein expression at low temperature (15°C), mild flash freezing and thawing to lyse cells, purification under denatured conditions and renaturing protein on Ni-NTA column, and protease inhibitor cocktail (Roche). These strategies, as described in Materials and Methods, gave

reasonable yields of about 10 mg purified protein per liter *E. coli* cells and significantly countered degradation during expression and purification, as shown under the 0 min trypsin digestion lanes for each construct (Fig.4).

The naturally occurring osmolyte TMAO can induce cooperative folding transitions in the NTDs of different GR translational isoforms

With excitation at 295nm, tryptophan emission spectra were measured for GR's A-NTD, C2-NTD and C3-NTD in buffers with different TMAO concentrations. All three constructs contain two tryptophans corresponding to the tryptophans in full length NTD at residues 213 and 364. As shown in Fig. 3 inset, in absence of TMAO, the maximal emission wavelength is at 348 nm with excitation at 295 nm for GR A-NTD construct (the same case was observed for GR C2-NTD and GR C3-NTD constructs, data not shown), indicating the tryptophans are solvent exposed in these ID proteins. In presence of 2.2 M TMAO, the emission wavelength maxima show a significant blue shift (338 nm versus 348 nm, see Fig.3), indicating that TMAO induced a more ordered conformation in these disordered proteins. Fluorescence intensities at 338 nm as a function of TMAO concentration for each construct were fitted successfully to the cooperative two-state transition model using LEM [46] as shown in Fig. 3. Free energies, $\Delta G_{U \rightarrow F}$, representing the stability of the folded state with unfolded state as reference, were fitted to be 10.1 ± 0.9 , 8.9 ± 0.4 and 7.6 ± 0.4 kcal/mol for GR's A-NTD, C2-NTD and C3-NTD, respectively. M-values, representing the cooperativity of the transitions, were fitted to be 8.8 ± 0.4 , 6.9 ± 0.4 and 6.7 ± 0.3 kcal/ (mol*M) for GR A-NTD, GR C2-NTD and GR C3-NTD, respectively (as shown in Table 1).

Protease protection assay validates the rank order of stability estimates from TMAO folding of the GR A-NTD, GR C2-NTD, and GR C3-NTD constructs

The stability measurements obtained from TMAO-induced folding provided an estimate of the stability of the folded state. To determine whether the apparent free energies from Table 1 are reporting on the relative stabilities of the unfolded proteins in the absence of TMAO, limited trypsin digestions of GR's A-NTD, C2-NTD, and C3-NTD were performed (as shown in Fig. 4). One hallmark of ID proteins is their high inherent sensitivity to protease digestion [48]. As expected, all three constructs are highly sensitive to trypsin. With a protein versus trypsin mass ratio 1000:1 at 22°C, after 5min, significant degradation could be observed for all constructs, and after 30 min, all approached complete digestion. Comparing the protease sensitivity, the A-NTD is the most sensitive while C3-NTD is the most stable. This observation is consistent with the stabilities of these constructs extrapolated from TMAO folding experiments, GR C3-NTD is ~1.3 kcal/mol more stable than C2-NTD, and ~2.5 kcal/mol more stable than A-NTD. Both the extrapolated free energy values and sensitivities to trypsin digestion for A-NTD, C2-NTD and C3-NTD indicate that truncation of the N-terminus of GR stabilize this ID domain.

The folded states of GR NTDs induced by TMAO are thermodynamically similar to globular proteins of similar length

One important thermodynamic parameter from the application of LEM [46] to osmolyte induced protein folding is the m-value, which is defined as follows [46]: $\Delta G_{u \rightarrow F}(TMAO) = \Delta G_{u \rightarrow F}^0 - m \times C_{TMAO}$. The m-value, which has the unit of kcal/(mol*M), represents the molar efficacy of osmolyte in folding a protein, and also measures the cooperativity of the transition between the unfolded and folded state [49]. For a cooperative two-state transition induced by a particular osmolyte, the m-value is proportional to the surface area buried upon protein folding, which, in turn, is generally proportional to the protein size [49]. The linear

relationship between m-values and globular protein chain lengths have been reported for denaturant induced protein unfolding [50]. In Fig. 5, the reported m-values between the fully unfolded and fully folded states for TMAO folding, Barstar [51], RCAM-T1 [37], P protein [52] and Nank1-7 [53], are shown as a function of protein length. For comparison, m-values of TMAO induced folding of GR A-NTD, GR C2-NTD and GR C3-NTD determined in this study are also depicted. These seven data points could be fitted to a line with correlation coefficient $R^2=0.97$. Two illustrations can be pointed out from this observation. First, it indicates that GR's A-NTD, C2-NTD and C3-NTD undergo cooperative transitions with increasing TMAO concentration, only the fully unfolded state and the fully folded state are populated during the transition. Second, the m-values for the long GR N-terminal domains and m-values for the globular proteins Barstar, RCAM-T1, P protein, Nank 1-7, (which are in the unfolded state under physiological conditions by destabilizing mutations) could be linearly fitted, demonstrating that full-length NTD and the two truncated translational isoforms are able to adopt native folds that are thermodynamically similar in terms of surface area burial to those adopted by globular protein. However, although seemingly unlikely, we cannot exclude the possibility that both the folded and unfolded states of ID and globular proteins are different from each other and respond differently to TMAO but that the differences are quantitatively canceled out in the overall m-value.

Full length GR NTD contains at least two thermodynamically coupled regions

The transactivation domain AF1, which is mapped to residues 77-262 within NTD in human GR, is indispensable for its maximal transcriptional activity [40]. However, the roles of the other domains in the NTD are not clear. As the activities of GR isoforms are acutely affected by truncation from the N terminus of AF1 [25], it suggests that the fragment containing the first 97

amino acids in the GR NTD acts as a “regulatory (R) domain or domain”. To check whether the R and F domains are coupled to each other, an ensemble allosteric model (EAM) [6] was constructed for the full length GR NTD considering that it is composed of R and F domains (as shown in Fig. 6). To quantify the effect, we needed to obtain estimates of the intrinsic stabilities and interaction energies (i.e., ΔG_R , ΔG_F and Δg_{int}). We note that the stability of the F domain can be directly obtained from the TMAO-induced folding of the C3 construct, $\Delta G_F = -\Delta G_{U \rightarrow F}^0$ (GR C3-NTD) = -7.6 kcal/mol. Determination of the stability of the R domain is less straightforward. The small size (97 a.a.) and the lack of Trp residues in this region make it impossible to observe and thermodynamically characterize with TMAO (as done with the F domain). Nonetheless, the CD signal of the GR1-97 construct (in the absence of TMAO) provides an important clue about the range of possible stabilities of the R domain.

As shown in Fig. 7, the relative CD signals at 200 and 222 nm for the GR 1-97 construct places it somewhere between the average signal for folded and ID proteins [54, 55]. This suggests that both folded and unfolded states must be populated to a significant degree. As only a narrow range of stability allows for both folded and unfolded species to co-exist in substantial quantities simultaneously (i.e., $\sim -1.5 < \Delta G_R < \sim 1.5$ kcal/mol), to a first approximation the folded and unfolded states can be estimated to be equally probable (i.e., $\Delta G_R \sim 0$ kcal/mol). Using these estimates for the stabilities of the F and R domains (i.e., $\Delta G_F = -7.6$ kcal/mol and $\Delta G_R = 0$ kcal/mol are fixed), the data for the TMAO dependence of the fluorescence of GR A-NTD can be fitted to obtain an estimate of the interaction energy.

The fitted value of Δg_{int} in the absence of TMAO ($\Delta g_{int}^0 = -3.0 \pm 1.0$ kcal/mol) indicates that the F and R domains are negatively coupled, and that the parameters are able to accurately

reproduce both the GR A-NTD and GR C3-NTD TMAO refolding curves (Fig. 8). We also note that the results are not qualitatively sensitive to the precise value of ΔG_R ; systematically varying the magnitude between -1.0 and +1.5 kcal/mol nonetheless results in a negative Δg_{int} .

2.5 Conclusions

In conclusion, long NTDs in three GR translational isoforms were found to cooperatively fold into globular protein-like folded states in the presence of TMAO. The 420 amino acid long NTD of GR contains at least two thermodynamically distinct regions, the regulatory (R) domain and the functional (F) domain, which are negatively coupled to each other. This negative coupling is foundational to the allosteric mechanism we describe in the following chapters.

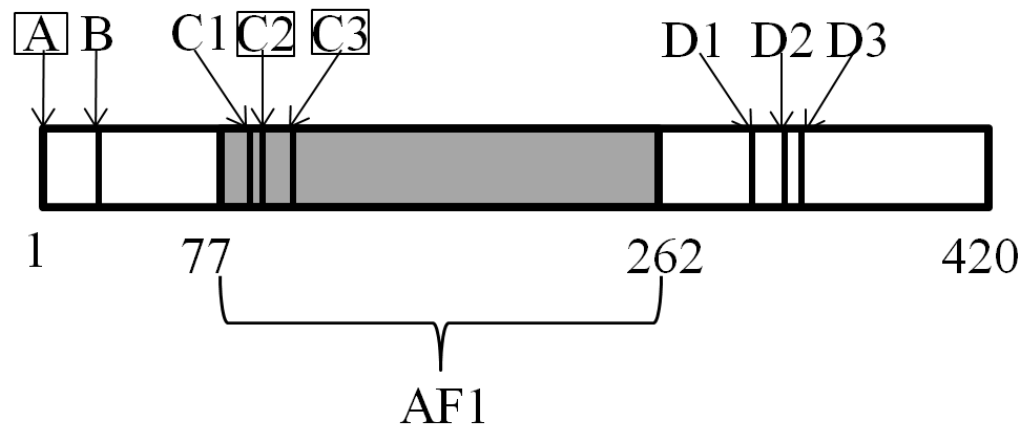


Figure 2 | Schematic representation of the GR's NTD sequence showing the positions of different GR translational isoform start sites.

The three constructs, A-NTD, C2-NTD and C3-NTD, which are thermodynamically characterized in this study, are boxed. The transactivation domain AF1 (GR 77–262) is shown in gray.

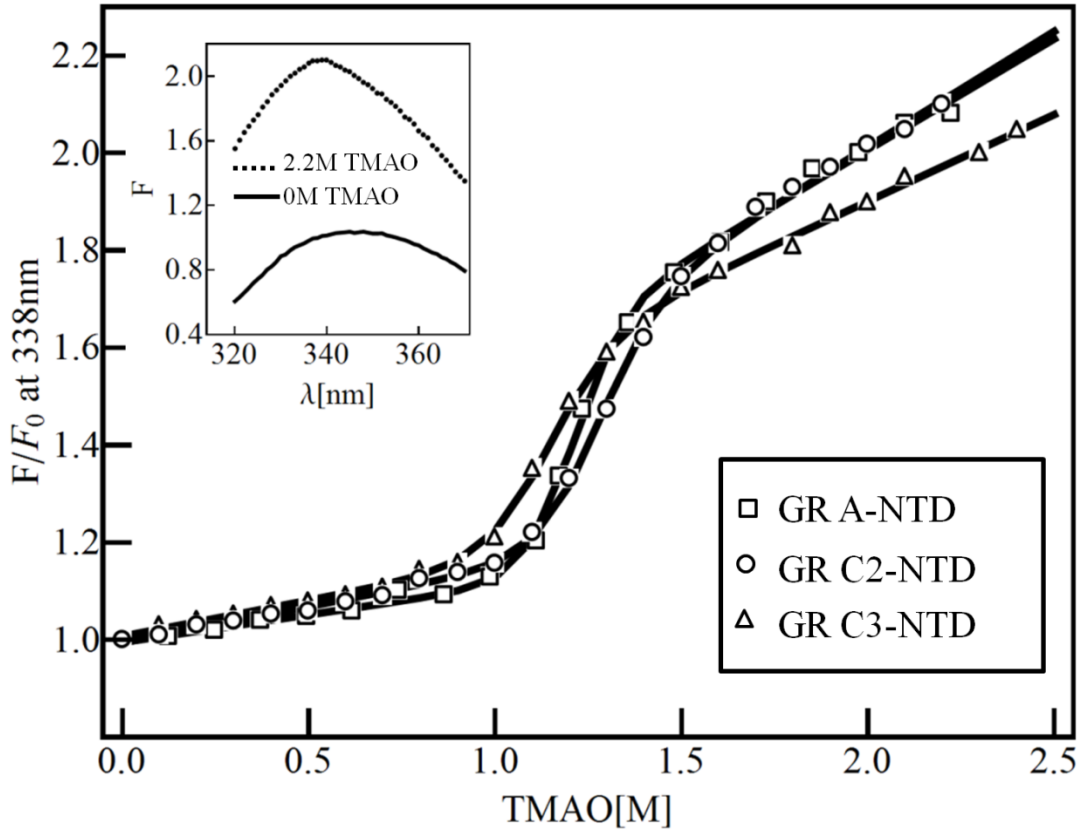


Figure 3 | TMAO-induced conformational transitions of GR A-NTD, C2-NTD, and C3-NTD monitored by tryptophan emission fluorescence intensity.

TMAO-induced conformational transitions of GR A-NTD (open squares), C2-NTD (open circles), and C3-NTD (open triangles) monitored by tryptophan emission fluorescence intensity at 338 nm with excitation at 295 nm. Fluorescence emission intensity at each TMAO concentration was normalized to that at 0 M TMAO concentration. The folding experiments were performed at 22 °C with a protein concentration of 0.5 μ M. Nonlinear least square fits to two-state transitions using the linear extrapolation method were carried out for each construct (solid line). The best-fit thermodynamic parameters are shown in Table 1. Inset, GR A-NTD fluorescence emission spectrum with excitation at 295nm in 0 M TMAO with λ_{max} at 348nm (solid line) and in 2.2 M TMAO with λ_{max} at 338 nm (dashed line).

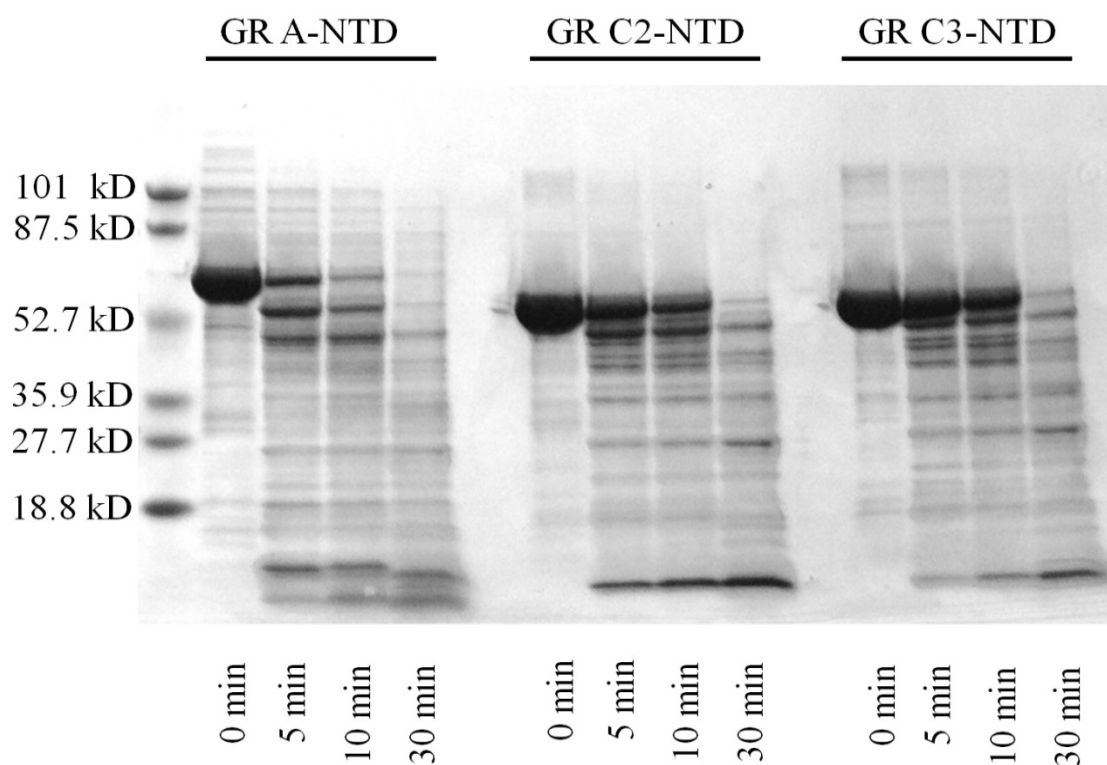


Figure 4 | Protease protection assays for GR A-NTD, C2-NTD, and C3-NTD.

Trypsin digestions were performed at a protein (1 mg/ml):trypsin mass ratio of 1000:1 at 22 °C in 10 mM HEPES, 80 mM NaCl, 1 mM EDTA, 10% glycerol buffer, pH 7.6 for 0, 5, 10, and 30 min. After quenching, 45 µg samples from each time point were separated on 4–15% SDS-PAGE gels with SDS-Tris-glycine gel running buffer.

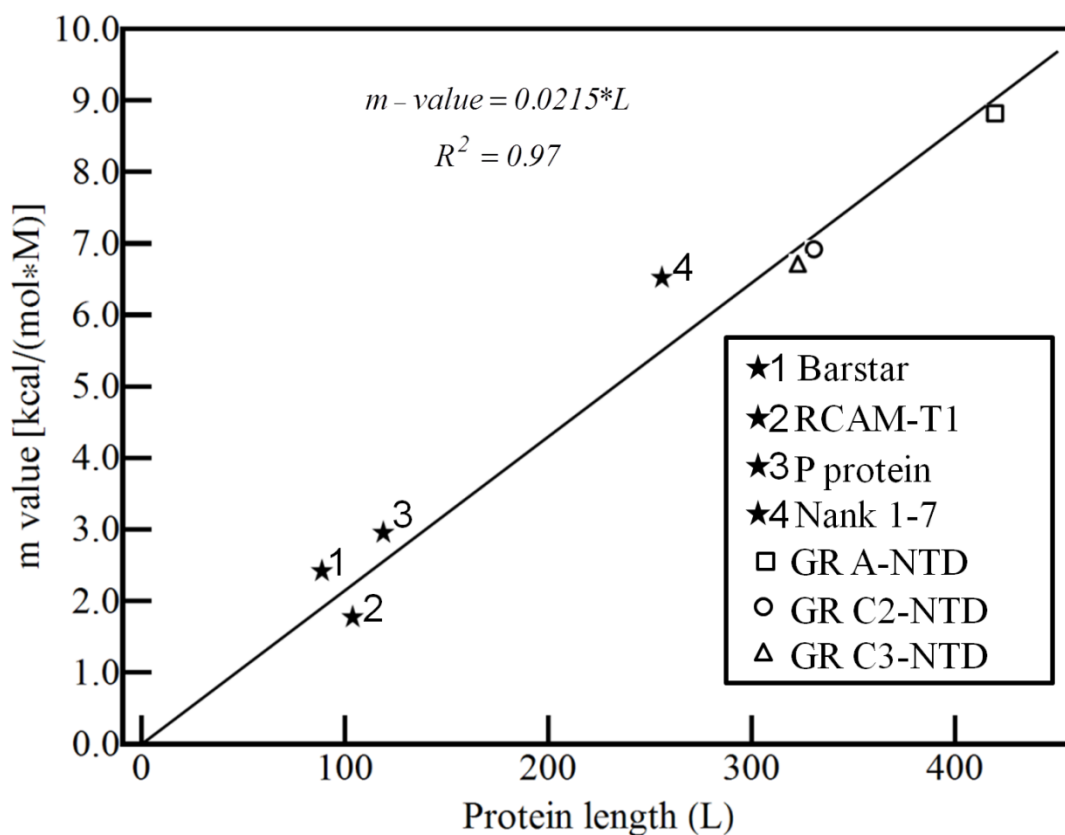


Figure 5 | Dependence of m-values on protein chain length (L) for TMAO induced folding transitions.

Shown are the m-values fitted from TMAO-induced two-state folding transitions for Barstar (closed star 1), RCAM-T1 (closed star 2), P protein (closed star 3), Nank 1–7 (closed star 4), GR's A-NTD (open square), C2-NTD (open circle), and C3-NTD (open triangle). All seven data points can be well fit to a line; i.e. $m\text{-value} = 0.0215 \cdot L$ with a correlation coefficient of $R^2 = 0.97$. This indicates that the TMAO-induced folding of GR A-NTD, C2-NTD, and C3-NTD followed a cooperatively two-state transition process. In addition, the folded states of GR NTDs induced by TMAO are thermodynamically similar to globular proteins of similar length.




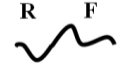
States		ΔG_i	$S.W. = e^{-\Delta G_i/RT}$	Probability
N		0	1	$P_N = \frac{1}{Q}$
I ₁		$\Delta G_F + \Delta g_{int}$	$e^{-(\Delta G_F + \Delta g_{int})/RT}$	$P_{I_1} = \frac{e^{-(\Delta G_F + \Delta g_{int})/RT}}{Q}$
I ₂		$\Delta G_R + \Delta g_{int}$	$e^{-(\Delta G_R + \Delta g_{int})/RT}$	$P_{I_2} = \frac{e^{-(\Delta G_R + \Delta g_{int})/RT}}{Q}$
U		$\Delta G_U = \Delta G_F + \Delta G_R + \Delta g_{int}$	$e^{-(\Delta G_F + \Delta G_R + \Delta g_{int})/RT}$	$P_U = \frac{e^{-(\Delta G_F + \Delta G_R + \Delta g_{int})/RT}}{Q}$

Figure 6 | Schematic representation of the ensemble allosteric model of full-length GR NTD containing the R domain and F domain.

Each region can be folded or unfolded, resulting in four possible states (i.e., N, I₁, I₂ and U). S.W. is the statistical weight for each state, and Q is the partition for the system, which is the sum of the statistical weights for all the states.

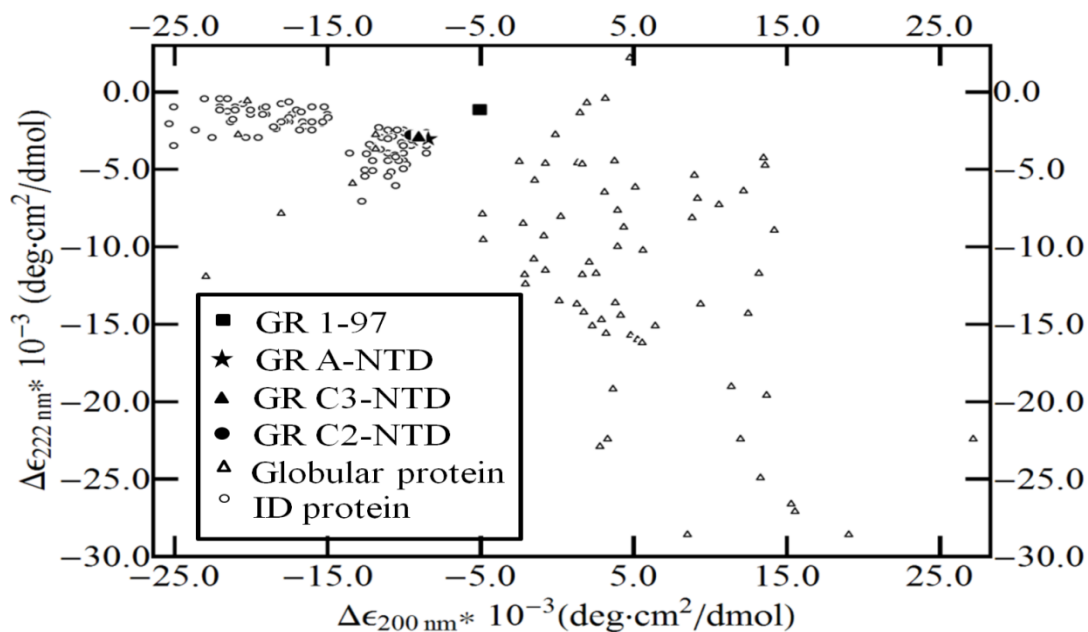


Figure 7 | Circular dichroism data comparison for globular and ID protein in terms of $\Delta\epsilon_{222 \text{ nm}}$ versus $\Delta\epsilon_{200 \text{ nm}}$

Open triangles represent the data from the Protein Circular Dichroism Data Bank [54, 55], and open circles represent the data for disordered protein [54, 55]. The filled star, triangle, and circle represent the data for GR A-NTD, C2-NTD, and C3-NTD, respectively, measured in this study. The filled square represents the data for GR 1-97, which is midway between the clusters for ID and globular folded proteins. deg, degrees. This supports the conclusion that the R domain in isolation is marginally stable even in the absence of TMAO.

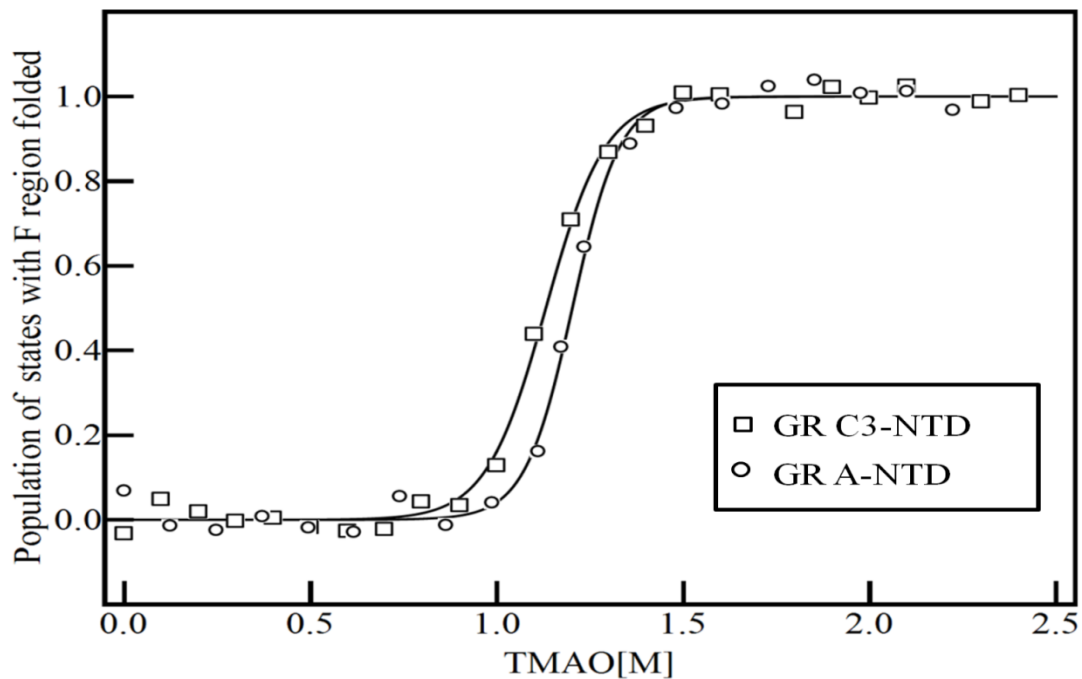


Figure 8 | Dependence of the population of the states with F domain folded on TMAO concentration for A-NTD and C3-NTD.

Raw data for GR A-NTD (open squares) and C3-NTD (open triangles) are calculated from the TMAO folding curve as shown in Fig. 3. The underlying assumption made is that the fully unfolded state U and the intermediate state I₁ (where the F domain is unfolded) have the same fluorescence emission intensity (set to 0) and that the fully folded state N and the intermediate state I₂ (wherein the F domain is folded) have the same fluorescence emission intensity (set to 1) as both of the two NTD tryptophans are in the F domain. The fitted curve is based on the linear extrapolation method, fitting to the equation that observable equals the sum the probability of the N state and I₂ state as shown in Fig. 6. The fitted parameters are as follows: $\Delta G_{F \rightarrow U}^0(F) = -8.4 \pm 0.5$ kcal/mol, $m(F) = 7.4 \pm 0.4$ kcal/(mol*M); $\Delta G_{F \rightarrow U}^0(R) = 0$ kcal/mol, $m(R) = 2.1 \pm 0.4$ kcal/(mol*M); $\Delta g_{int}^0 = -3.0 \pm 1.0$ kcal/mol, $m_{int} = 2.0 \pm 0.9$ kcal/(mol*M).

Table1. Fitting parameters for GR A-NTD, GR C2-NTD and GR C3-NTD constructs from TMAO induced folding transitions.

Constructs	$\Delta G_{U \rightarrow F}$ (kcal/mol)	m-value (kcal/(mol*M))
GR A-NTD	10.1 \pm 0.5	8.8 \pm 0.4
GR C2-NTD	8.9 \pm 0.4	6.9 \pm 0.3
GR C3-NTD	7.6 \pm 0.3	6.7 \pm 0.3

Chapter 3- Thermodynamic Dissection of the NTD and DBD

Two-domain Constructs of Human Glucocorticoid Receptor

3.1 Abstract

The existence of inter-domain allosteric coupling between NTD and DBD in GR has been reported in the literature; however, there is no reported quantitative investigation. As TMAO induced folding transitions can be used to quantify the thermodynamic parameters of the NTD of GR isoforms as discussed in Chapter 2, we also carried out these folding experiments on the GR A-NTDDBD and C3-NTDDBD constructs. These studies were aimed at quantifying the thermodynamic coupling between the domains. As expected, these two-domain constructs were also foldable using TMAO.

However, judging from the *m*-value, which provides insight into the protein folding/unfolding cooperativity, GR A-NTDDBD (composed of R, F and DBD as shown in Fig. 1) did not follow a cooperative, two-state folding transition and only C3-NTDDBD (composed of F and DBD as shown in Fig. 1) obeyed the cooperative, two-state folding transition. Comparing the stability extrapolated for C3-NTD and C3-NTDDBD constructs, we found that the NTD of C3 was stabilized in presence of DBD. This indicates there is favorable coupling between the F domain in the ID NTD and the DBD.

3.2 Introduction

One characteristic of proteins in the SHR family is their domain architecture. Most of the members contain three major domains, the intrinsically disordered N-terminal domain (NTD), the DNA binding domain (DBD) and the ligand binding domain (LBD) [15]. The LBD contains a binding site for hormones, the DBD contains a binding site for the hormone response element

(HRE), and the NTD contains the major transactivation region, AF1, which is the binding site for most co-regulators. Coupled folding and binding have been suggested to be the key for the interaction of the NTD and its co-regulators [15]. Clearly, linking multiple domains with different functions in steroid hormone receptors make them work as hormone dependent nuclear transcription factors, as they are only functional in the complex with hormone, HRE and co-regulators bound. Except for executing their individual functions, do individual domains also communicate with neighboring domains and influence each other's functions? The strategy to dig into the biophysical origins of the functional communications is to ask whether neighboring domains influence each other's thermodynamics or folding kinetics [56].

In GR, the inter-domain allosteric coupling between NTD and DBD has been reported in the literature [18, 19, 22]. For instance, R.Kumar and colleagues found that binding of DBD to the GR response element (GRE), the DNA sequence that could bind to the DBD, induced additional structure in the NTD [18]. In addition, Yamamoto's group found that mutations within the activation function 1 (AF1) region in NTD exhibited different influences on GR transcriptional activity when bound to different GRE [19]. These two examples demonstrate that there is significant energetic coupling between GR NTD and DBD, and this coupling influences GR's function as a transcription factor. However, so far there is no reasonable quantification of the allosteric coupling.

To further study the inter-domain coupling between NTD and DBD, we carried out TMAO induced folding transitions on GR two-domain constructs containing NTD and DBD. As expected, the folding transitions of both GR A-NTD and C3-NTD are influenced by the presence of DBD. Specifically, GR A-NTDDBD construct, composed of R domain, F domain and DBD, did not

follow cooperative two-state transitions. GR C3-NTDDBD construct, composed of F and DBD, turned out to follow cooperative two-state transitions. And the fitted parameter suggested that the F domain is stabilized in the presence of the DBD. In both cases, DBD influences the folding of its neighboring NTD, which suggests that there are domain interfaces between the folded state of NTD and DBD [56].

3.3 Materials and methods

Plasmids

Codons encoding human GR A-NTDDBD were optimized for *E. coli* expression with 6 histidines tagged on the N terminus and synthesized by DNA 2.0 (Menlo Park, CA), and inserted into the PJ411 bacterial cell expression vector under T7 promoter control. Plasmid PJ411-C3-NTDDBD was made by inserting the codons for C3-NTDDBD amplified from PJ411 A-NTDDBD into the NdeI and XhoI sites of the PJ411 vector.

Protein expression, purification and storage

Protocols for expression, purification and storage of the two-domain constructs were the same as the protocols for the single N terminal domain construct as described in chapter 2, except for the following modifications in the purification steps on Ni affinity column. The lysis buffer is composed of 100mM monosodium phosphate, 10mM Tris, 500mM NaCl, 20mM imidazole, pH 8.0. The wash buffer is the lysis buffer containing 60mM imidazole and the elution buffer is the lysis buffer containing 200mM imidazole.

Osmolyte TMAO induced protein folding transitions

Protocol for TMAO induced folding transitions is the same as described in chapter 2.

Protease protection assay

The protease protection assay was carried out in the same way as done for the single NTD constructs as describe in chapter 2, except for the following modification. The protein and trypsin ratio was 500:1 instead of 1000:1 in single NTD constructs' digestion.

3.4 Results and discussion

TMAO induced folding of GR A-NTDDBD construct does not follow cooperative two-state transition

Tryptophan emission spectra were measured for GR A-NTDDBD construct in buffers with different TMAO concentrations with excitation at 295nm. The construct contains two tryptophans corresponding to the tryptophans in full length NTD at residues 213 and 364. As shown in Fig. 9, in absence of TMAO, the maximal emission wavelength was at 348nm for GR A-NTDDBD construct, indicating the tryptophans are solvent exposed in this ID region. In the presence of 2.5M TMAO, the emission wavelength maximum shows a significant blue shift (338nm versus 348nm), indicating that TMAO induced a more ordered conformation in this disordered protein. As the sigmoid curve was observed, normalized fluorescence intensities at 338nm were fitted to the cooperative two-state transition model using LEM [46] as shown in Fig. 10. Free energy, $\Delta G_{U \rightarrow F}$, representing the stability of the folded state with unfolded state as a reference, was fitted to be 5.7 kcal/mol for GR A-NTDDBD. The m-value, representing the cooperativity of the transition, was fitted to be 4.9 kcal/(mol*M) as listed in Table 2. Noticeably from Fig. 10, the slope of the A-NTDDBD folding transition is less than that for the A-NTD construct. As A-NTDDBD and A-NTD constructs have the same length of the intrinsically disordered NTD (420 amino acids), and both of the two tryptophans monitored in the fluorescence spectrum are in the NTD, the significantly different folding transitions between A-NTDDBD and A-NTD indicate

that DBD influences NTD's folding properties. As shown in Fig. 2 and discussed in chapter 2, A-NTDDBD construct is composed of the R domain, F domain and DBD. Thus there may be some interaction interface either between the R domain and DBD, or between the F domain and DBD, or both of these cases.

If mapping the m-value for the A-NTDDBD construct to the graph of protein chain length and expected m-value, as shown in Fig. 12 we can see that the m-value for A-NTDDBD is much smaller than expected for the 420aa long construct following two-state folding transition in TMAO. This indicates that GR A-NTDDBD does not follow a cooperative two-state transition, and there may be some intermediate states populated during the transition, even though a sigmoid curve was observed. So the free energy fitted is the apparent free energy of the unfolded to folded transition, it does not truly represent the free energy difference between the folded and unfolded state.

TMAO could induce cooperative folding transitions in C3-NTDDBD construct and the folded conformation of its ID part is stabilized in presence of DBD

As done for A-NTDDBD, the C3-NTDDBD emission spectrum was also recorded with excitation at 295nm under different TMAO concentrations. Similarly, the emission wavelength maximum showed a significant blue shift with increasing TMAO concentrations, which indicated that the construct adopted a more compact conformation in TMAO. Emission intensity at 338nm was recorded as a function of TMAO concentration, and the distance to the two baselines was plotted as a function of TMAO concentration as shown in Fig. 11. This folding curve was fitted to the cooperative two-state transition model using LEM [46]. Free energy, $\Delta G_{U \rightarrow F}$, representing the stability of the folded state with the unfolded state as reference, was fitted to be 6.7 ± 0.4 kcal/mol

for GR C3-NTDDBD. The m-value was fitted to be $7.0 \pm 0.5 \text{ kcal/ (mol*M)}$ as listed in Table 2.

If mapping the m-value for the C3-NTDDBD construct to the relationship between protein chain length and the expected m-value curve for the cooperative two-state transition as shown in Figure 11, we can see that the m-value for C3-NTDDBD is only a little bit larger than the expected m-value for C3-NTD. This indicates that GR C3-NTDDBD (composed of the F domain and DBD) probably follows a cooperative two-state transition in TMAO. And the interaction interface between folded conformation of F and DBD contributes some to the m-value. The fitted free energy, $\Delta G_{U \rightarrow F}$, can be deconvoluted to the free energy difference between the folded state and unfolded state of the F domain, plus the free energy difference between the low affinity state and high affinity state of DBD to GRE, and the interaction energy between F domain and DBD. The derived equations are shown below with the folded state as a reference:

$$\Delta G_{F \rightarrow U} = -\Delta G_{U \rightarrow F} = \Delta G_{F \rightarrow U} (F) + \Delta G_{H \rightarrow L} (DBD) + \Delta g_{\text{int}} (F, DBD) = -6.7 \text{ kcal/mol}$$

Where $\Delta G_{F \rightarrow U} (F) = -7.6 \text{ kcal/mol}$, extrapolated from the TMAO induced folding transition of the C3-NTD construct in Chapter 2, so $\Delta g_{\text{int}} (F, DBD) = 0.9 - \Delta G_{H \rightarrow L} (DBD)$. Comparing the DBD's crystal structure in presence of GRE to its NMR solution structure in absence of GRE, the second and third helices are stabilized upon GRE binding. This suggests that in solution the more favorable state for DBD is the low affinity state, thus $\Delta G_{H \rightarrow L} (DBD) < 0$. Consequently, $\Delta g_{\text{int}} (F, DBD) = 0.9 - \Delta G_{H \rightarrow L} (DBD) > 0$. This indicates that there is a favorable interaction between the folded conformation of the F domain and the DBD.

Protease Protection Assay suggests allosteric coupling between R domain and DBD

The protease protection assay was also carried out for GR A-NTDDBD construct and GR C3-NTDDBD construct at 22°C with a protein vs trypsin ratio of 500 vs 1. As shown in Fig. 13,

GR's A-NTDDBD is significantly protected from trypsin digestion compared to C3-NTDDBD and A-NTD, as shown in Fig. 4. This suggests that there may be a significant interaction between the R domain and DBD, and this interaction protects the NTD from protease digestion. This hypothesis will be further tested in the following chapters.

3.5 Conclusions

GR A-NTDDBD, composed of the R domain, F domain and DBD, did not follow a cooperative, two-state transition in TMAO. GR C3-NTDDBD, composed of the F domain and DBD, followed a cooperative, two-state transition in TMAO. The folded conformation of the F domain is stabilized in the presence of the DBD, which suggests there is favorable interaction energy between the F domain and DBD. Moreover, the protease protection assay comparing GR A-NTDDBD, C3-NTDDBD and A-NTD suggested that there may be some interaction between the R domain and DBD, which provided us a hypothesis to test in the following chapters.

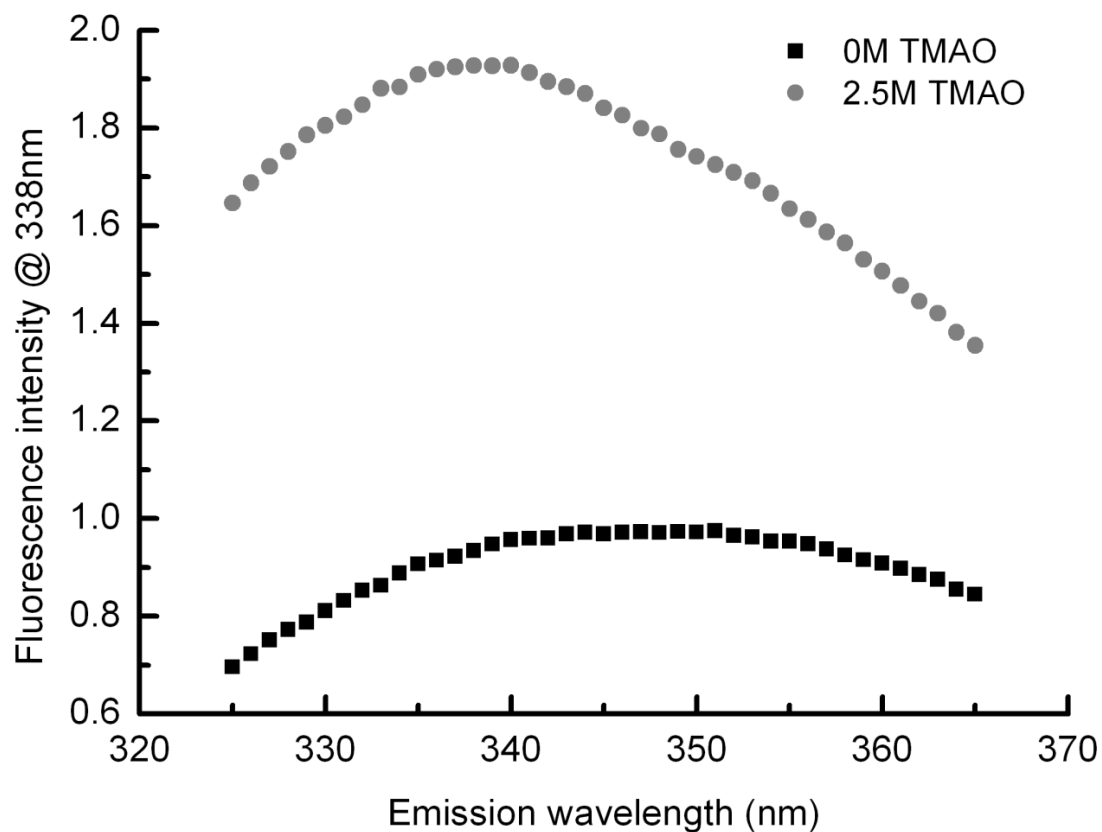


Figure 9 | Fluorescence emission spectrum of GR A-NTDDBD at 0M TMAO vs. 2.5 M TMAO.

At 0 M TMAO the emission spectrum maximum wavelength was at 348nm, and 2.5 M TMAO it was blue shifted to 338nm. The blue shift suggests that the ID region adopted a more ordered conformation in the presence of 2.5 M TMAO.

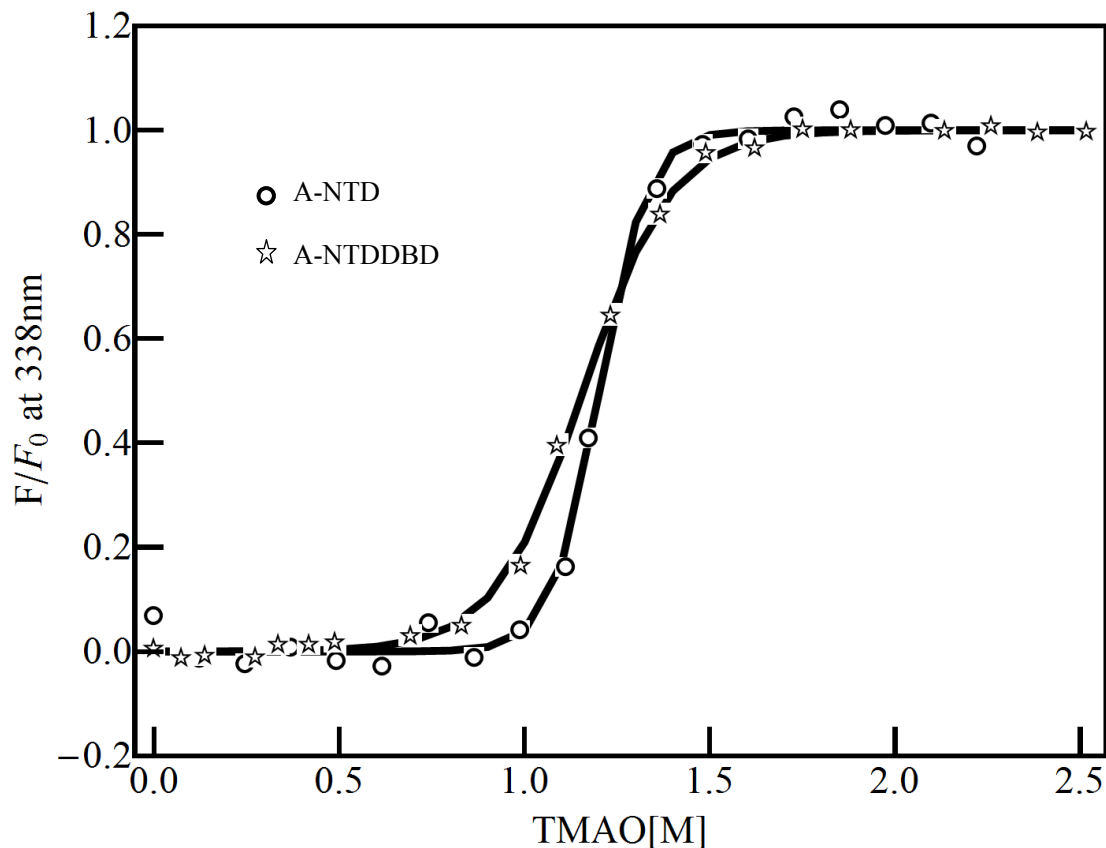


Figure 10 | Comparison of TMAO-induced conformational transitions of GR A-NTDDBD and GR A-NTD monitored by tryptophan emission fluorescence intensity.

TMAO was used to fold GR's A-NTDDBD and A-NTD, monitored by tryptophan emission intensity with excitation at 295nm. Fluorescence intensity at 338nm was converted to the folded state probability, by calculating the distance of the raw signal to the unfolded and folded state baseline. The curve shown is the raw data fitted with LEM [46]. The best fit thermodynamic parameters are shown in Table 2.

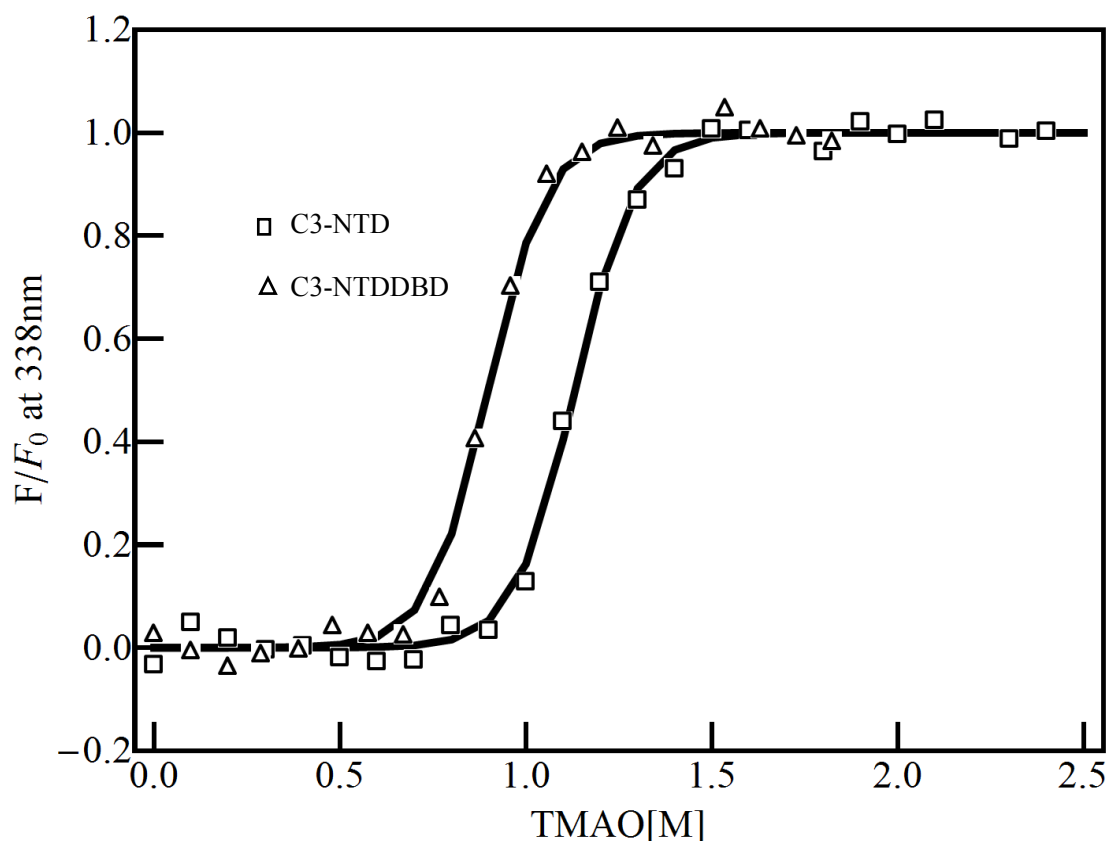


Figure 11 | TMAO-induced conformational transitions of GR C3-NTDDBD and GR C3-NTD monitored by tryptophan emission fluorescence intensity.

TMAO was used to fold GR's C3-NTDDBD and C3-NTD, monitored by tryptophan emission intensity with excitation at 295nm. Fluorescence intensity at 338nm was converted to the folded state probability, by calculating the distance of the raw signal to the unfolded and folded state baseline.. The curve shown is the raw data fitted with LEM [46]. The best fit thermodynamic parameters are shown in Table 2.

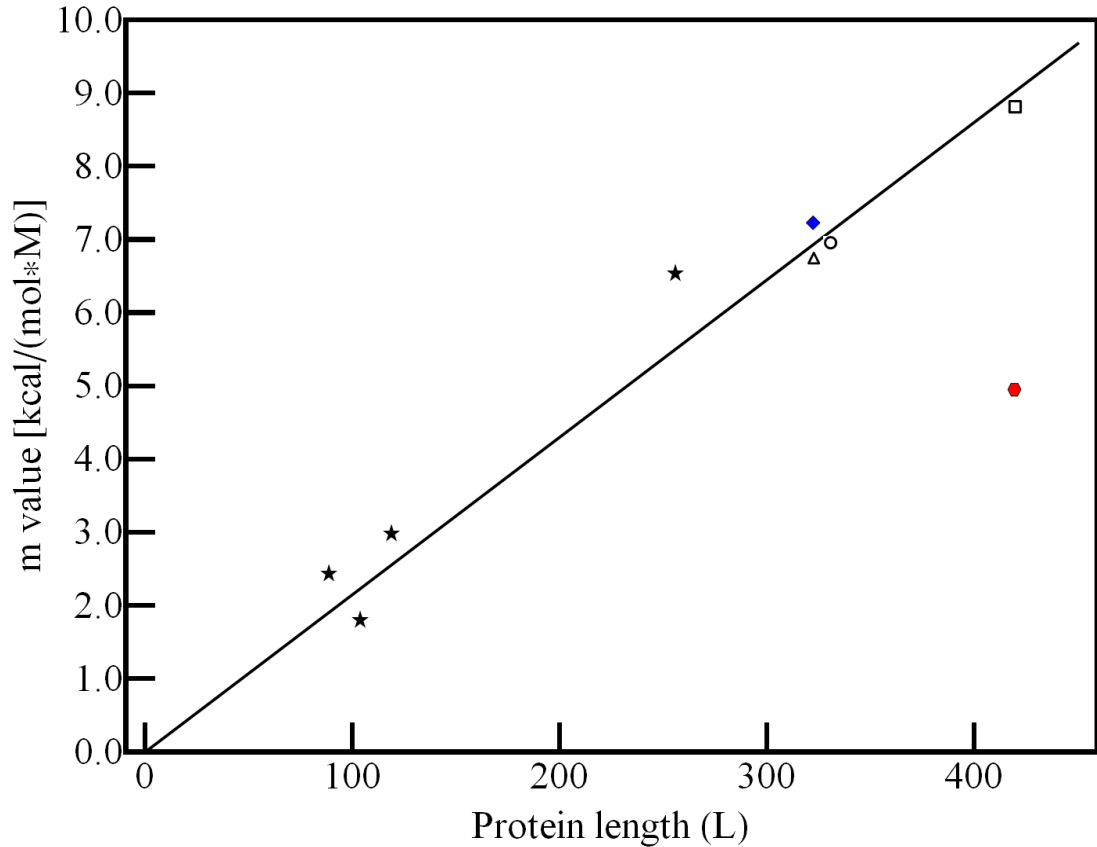


Figure 12 | Mapping m-value for GR A-NTDDBD and GR C3-NTDDBD on the m-value versus protein chain length chart.

The m-value for GR A-NTDDBD significantly deviates from the expected value for a protein of that chain length, indicating GR A-NTDDBD doesn't follow a two-state folding transition in TMAO. The m-value fitted from a two-state folding transition for GR C3-NTDDBD is a little bit larger than the expected value for the protein of that chain length, which is reasonable as there may be some interaction interface between the N terminal domain and DNA binding domain contributing to the m-value. Symbols are identical to those in Fig.5, except for the filled red circle, which corresponds to GR A-NTDDBD, and the filled blue diamond, which corresponds to GR C3-NTDDBD.

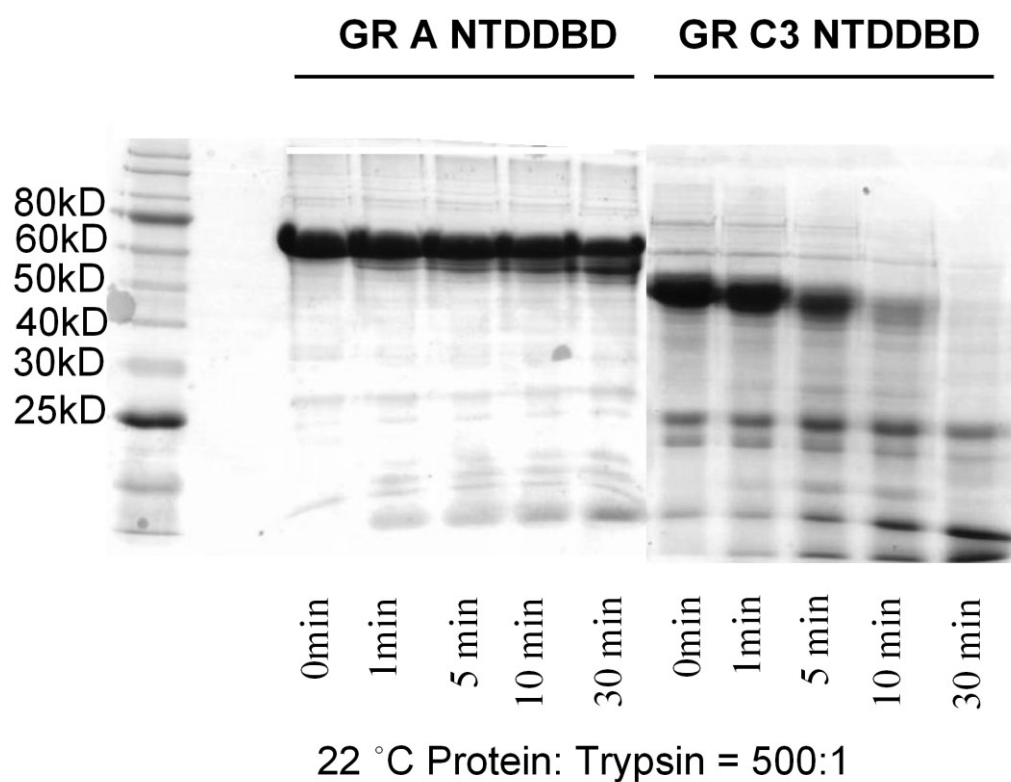


Figure 13 | Protease protection assays for GR A-NTDDBD and GR C3-NTDDBD.

Trypsin digestions were performed at a protein (1 mg/ml): trypsin mass ratio of 500:1 at 22 °C for 0, 1, 5, 10, and 30 min, respectively. After quenching, 45 µg samples from each time point were separated on 4–15% SDS-PAGE gels.

Table2. Fitted parameters for GR A-NTD, GR A-NTDDBD, GR C3-NTD, and GR C3-NTDDBD constructs from TMAO induced folding transitions.

Constructs	$\Delta G_{U \rightarrow F}$ (kcal/mol)	m-value (kcal/(mol*M))
A-NTD	10.1 \pm 0.5	-8.8 \pm 0.8
A-NTDDBD	5.7 \pm 0.4	-4.9 \pm 0.3
C3-NTD	7.6 \pm 0.3	-6.7 \pm 0.4
C3-NTDDBD	6.7 \pm 0.4	-7.0 \pm 0.5

Chapter 4- Binding Energetics of the Two-domain Constructs of GR

Translational isoforms to GR Response Element

4.1 Abstract

Allostery, by which remote sites of a protein are energetically coupled to elicit a functional response, manifests as action at distance. Thus exploring one region's influence on the coupled region's binding energetics to its ligands will help to understand the coupling mechanism. To gain further insight into the allosteric coupling between NTD and DBD, binding affinities to the half site of a GRE were measured for the NTDDDBD constructs of the eight GR translational isoforms. The binding was monitored by the change in the fluorescence anisotropy of the half site GRE labeled with fluorescent dye 6-FAM. The measured binding affinities differed about 5 fold, ranging from 8.4×10^5 to 4.2×10^6 , corresponding to the binding affinities of C2-NTDDDBD and D2-NTDDDBD, respectively. The length of R domain modulated the DBD's binding affinity to GRE, as seen in the A, B, C1 and C2 isoforms. This suggests that the R domain is coupled to the DBD, either directly or indirectly. The D2-NTDDDBD had the highest binding affinity, which suggests that the D2 region of the NTD contributes significantly to the favorable coupling between the F domain and the DBD.

4.2 Introduction

One common role of a DBD in transcription factors is to anchor the transcription factor to the response element on a chromosome. The DBD of GR tethers it to the glucocorticoid response element (GRE) in the promoter domain of a regulated gene. A typical consensus GRE is a palindromic repeat of 5'-AGAACA-3' half site separated by a three base pair spacer [16]. Each half site is a binding site for one GR monomer. Many natural response elements do not completely

match the perfect palindromic consensus, and the usage frequency of each nucleotide at each position is reported and shown in Fig. 14 [22]. In addition, half-site GRE usually forms a composite response element with other transcription factors' response elements [57]. The transcriptional activity of GR is usually weak when bound to a half-site or a full length response element, and multiple response elements are often found in close vicinity in the regulatory region of GR controlled genes [16]. There is both inter-site cooperativity and intra-site cooperativity reported for GR [29]. In this chapter, the binding affinities to the typical half site GRE 5'-AGAACA-3' are measured for the two-domain constructs of the eight GR translational isoforms.

4.3 Materials and methods

Protein expression, purification and storage

Protein expression, purification and storage have been previously described in Chapter 3.

6FAM labeled half site GRE and control oligonucleotide

Oligonucleotides containing half site GRE (5'-6FAM gcgcAGAACAaggacgcg-3' and 5'-cgcgtccTGTTCTgcgc-3') were synthesized by Integrated DNA Technologies (IDT) with HPLC grade purification and annealed with each other to get the double stranded 6FAM labeled half site GRE. Nonspecific control oligonucleotides (5'-GCGCCATATGATACGCG-3' and 5' -CGCGTATCATATGGCGC-3') were synthesized as complimentary strands by IDT and annealed to each other to get the double stranded control oligonucleotide.

Binding affinity measurement monitored by fluorescence anisotropy change

The binding experiments were carried out in the following buffer: 10mM HEPES (pH7.6), 80mM NaCl, 1mM EDTA, 5mM MgCl₂, 1mM DTT, 200ug/mL BSA and 5uM control

oligonucleotide. For each data point, 25 nM 6-FAM labeled half site GRE was incubated with 0 μ M up to 10 μ M of GR two-domain construct at 22°C for 30 minutes. Fluorescence anisotropy for mixtures at each GR concentration was measured using an Aviv ATF 105 fluorometer equipped with polarizers. The GR construct and 6FAM labeled half site GRE mixture (150 μ L) was allowed to rest in a “sub micro” fluorometer cell at 22°C (Santa Cells) for 2 minutes to allow for temperature stabilization and then the mixture was excited at a wavelength of 495nm. Anisotropy at 521nm was recorded as a function of GR construct concentration.

Binding affinity fitted with single site binding model

Binding affinity was fitted with nonlinear least squares fitting to the raw data with the fitting equation as follows. Detailed derivation is in Appendix 1.

$$A_{obs} = (A_{boundGRE} - A_{freeGRE}) \frac{1 + KaC_{GRE} + KaC_{GR} - \sqrt{(1 + KaC_{GRE} + KaC_{GR})^2 - 4Ka^2C_{GRE}C_{GR}}}{2KaC_{GRE}} + A_{freeGRE}$$

Where A_{obs} is the observed anisotropy value at each GR concentration, $A_{freeGRE}$ is the anisotropy value of the free GRE oligonucleotide, $A_{boundGRE}$ is the anisotropy value of the GRE oligonucleotide in complex with GR, Ka is the binding affinity between the half site GRE and GR. The fitting parameters are $A_{freeGRE}$, $A_{boundGRE}$, and Ka .

4.4 Results and discussion

Eight GR translational isoforms have variable binding affinities to GRE

Binding affinities of the GR two-domain constructs to glucocorticoid response element (GRE) were measured by monitoring the change in anisotropy of the 6FAM labeled half-site GRE as shown in Figure 16. The fitted binding affinity for each isoform is shown as a bar plot in Fig. 17 and also listed in Table 3 with the other fitting parameters. The binding affinities for the eight translational isoforms varied approximately 5 fold. The two-domain constructs of the eight

translational isoforms all contain the same intact DBD to bind to the GRE. The variable binding affinity to the GRE indicates that the NTD of different translational isoforms allosterically modulates DBD binding affinity to GRE. The coupling between NTD and DBD is consistent with previous reports that the GRE is an allosteric effector of GR [16, 17, 19, 22]. Detailed examination of the binding affinities of the translational isoforms in light of their domain architecture (shown in Fig. 18) can suggest testable hypotheses.

For example, A, B, C1 and C2 isoforms all contain the DBD, the F domain and a variable sized R domain, which accounts for differences in binding affinities to GRE. We hypothesize that the R domain is allosterically coupled to the DBD, either directly or indirectly through the F domain.

In C3, D1, D2 and D3 isoforms, the R domain is entirely missing and they have variable lengths of the F domain. As discussed in chapter 3, the F domain is favorably coupled to DBD. The significantly higher binding affinity of D2 isoform suggests that the N terminal domain of D2 isoform may contribute significantly to the favorable coupling between F domain and DBD.

4.5 Conclusions

The NTDDBD constructs of the eight GR translational isoforms had variable binding affinities to the GRE. As they all have the same intact DBD, which contains the binding site to GRE, these results suggest that the NTD is allosterically coupled to the DBD, and the coupling is differentially modulated in different translational isoforms.

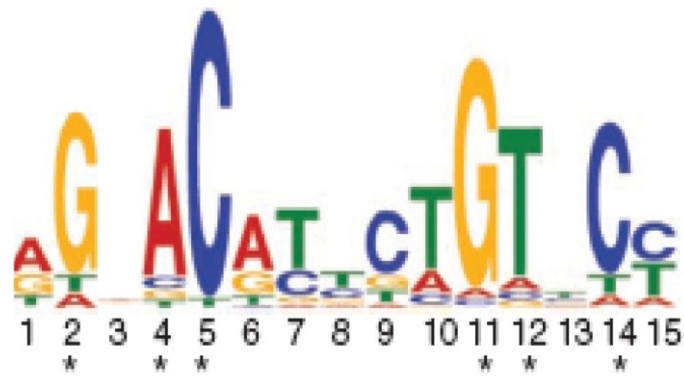


Figure 14 | Pattern of Glucocorticoid response element (GRE)

This is a figure modified from reference [22]. Positions 1-6 make up the left half-site, 10-15 make up the right half-site, and positions 7-9 makes up a spacer between the two half-sites. The two half-sites bind to GR DBD. The nucleotides that contact the DBD are labeled with asterisks.

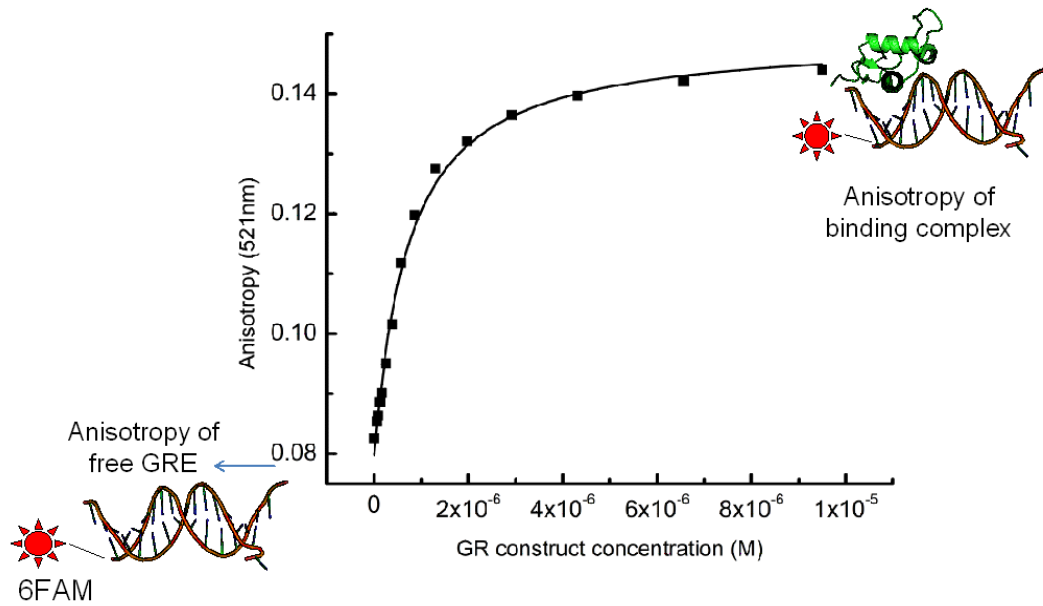


Figure 15 | Illustration of experimental design for binding GR's two-domain constructs to 6FAM labeled GRE.

In the binding experiment, 6FAM labeled GRE was kept at constant concentration. And concentration of the GR construct was gradually increased. In absence of GR, the GRE is in the free form and tumbling faster in the solution, so smaller fluorescence anisotropy is observed. In the presence of a saturating concentration of GR, the GRE is mostly in the bound complex and tumbles slower in solution (larger anisotropy). Monitoring the fluorescence anisotropy change with increasing protein concentration allows us to monitor the binding process.

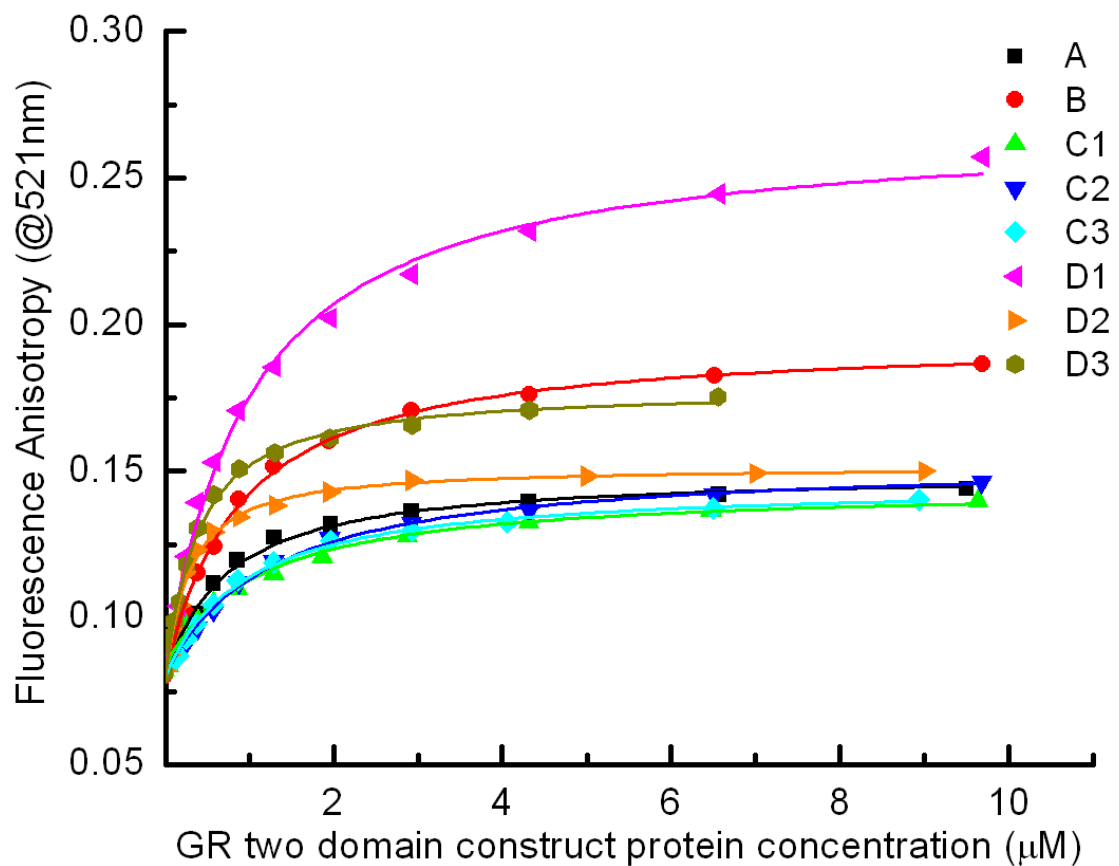


Figure 16 | Binding of NTDDDBD two-domain constructs of eight GR translational isoforms to GRE monitored by fluorescence anisotropy change

Fluorescence anisotropy change at 521nm of the 6FAM labeled GRE as a function of the concentration of GR isoforms. Binding was conducted with 25nM 6-FAM labeled GRE in buffer containing 10mM HEPES (pH7.6), 80mM NaCl, 1mM EDTA, 5mM MgCl₂, 1mM DTT, 10% glycerol, 200ug/mL BSA and 5uM control oligonucleotide. Curves were fitted with the single binding model and the best fitted parameters are shown in Table 3.

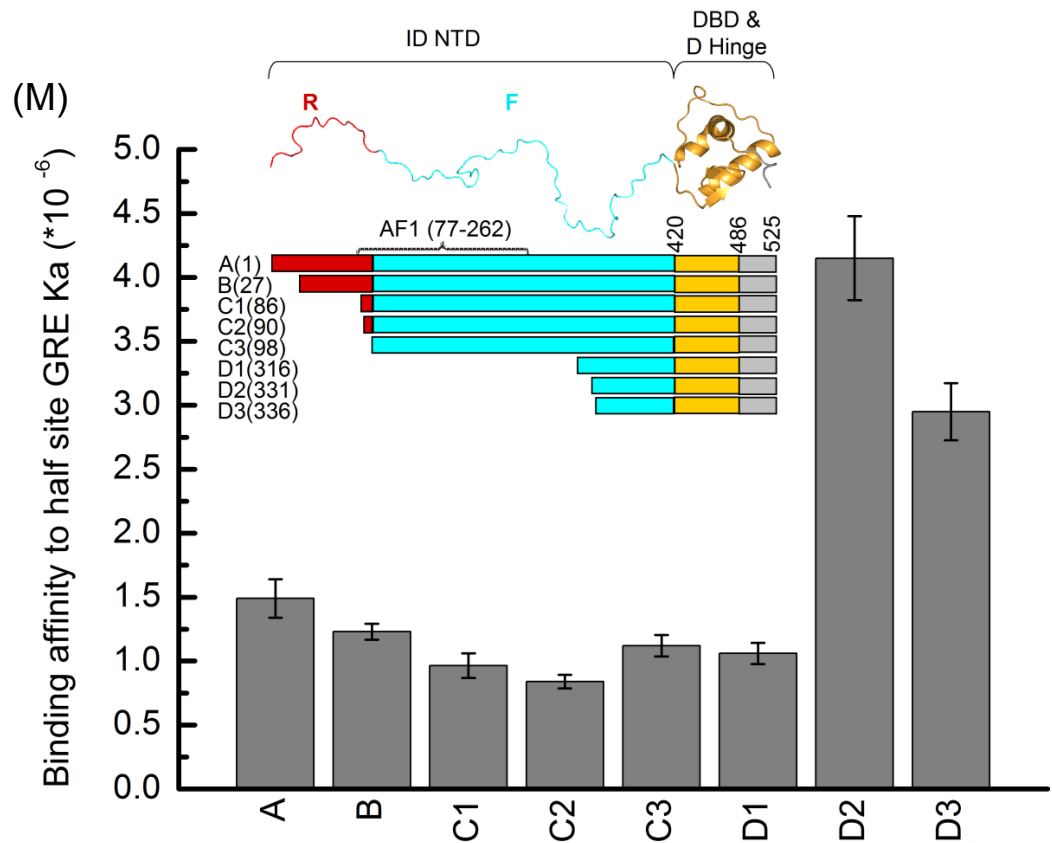


Figure 17 | Bar plot comparison of eight GR translational isoforms' binding affinity to half site GRE.

The binding affinity was fitted from the anisotropy data of Fig. 15. Shown in the inset is the domain structure of the eight, two-domain constructs used. Error bars indicate the fitting errors. Replicate measurements were within the fitting errors.

Table3. Fitted parameters for binding between NTDDDBD two-domain constructs of eight GR translational isoforms to 6FAM labeled GRE.

	Ka (M ⁻¹)		Ab		Af	
	Value	Fitting Error	Value	Error	Value	Error
A-NTDDDBD	1.49E+06	1.50E+05	0.14955	1.51E-03	0.07977	1.16E-03
B-NTDDDBD	1.23E+06	6.22E+04	0.19582	1.35E-03	0.07986	9.02E-04
C1-NTDDDBD	9.64E+05	9.64E+04	0.14501	1.56E-03	0.08026	8.71E-04
C2-NTDDDBD	8.40E+05	5.45E+04	0.15398	1.26E-03	0.08051	6.37E-04
C3-NTDDDBD	1.12E+06	8.37E+04	0.14593	1.18E-03	0.07974	7.33E-04
D1-NTDDDBD	1.06E+06	8.23E+04	0.26802	3.50E-03	0.08179	2.12E-03
D2-NTDDDBD	4.15E+06	3.29E+05	0.15176	8.94E-04	0.07935	1.19E-03
D3-NTDDDBD	2.95E+06	2.23E+05	0.1784	1.43E-03	0.07908	1.38E-03

Ab: fluorescence anisotropy of the bound complex

Af: fluorescence anisotropy of the free 6FAM labeled GRE

Chapter 5-Cell Based Functional Assay to Measure the Transcriptional Activity and Relative Binding Affinity of GR

5.1 Abstract

To further test the hypothesis generated from the biophysical studies, a dual luciferase reporter assay was developed using U-2 OS cells. Transcriptional activity at different GR expression vector dosages was measured for each translational isoform, and the maximum transcriptional activity and EC50 were fitted from these dosage curves. The EC50 of the active isoforms significantly correlated with the *in vitro* measured binding affinity from Chapter 4, indicating that the *in vitro* measured binding energetics represent the binding in cells. Different transcriptional activities of the eight translational isoforms were observed and did not correlate with their binding affinities to GRE. In addition, the luciferase assay was reformatted into a competitive transfection assay and shown to be valuable in measuring the *in vivo* binding strength of the inactive isoforms to GRE.

5.2 Introduction

The biological role of GR as a transcription factor provides a convenient way to assay its function through reporter gene assays. Development of a functional assay enables us to link the biophysical characteristics with protein function, and also to test the hypotheses generated from *in vitro* assays.

The different transcriptional activities of the full-length translational isoforms have been reported [25]. As the experimental system in this thesis are NTDDBD constructs of GR, it was necessary to assay the activity of these two-domain constructs. Truncation of the LBD removes the transcriptional activity dependence on hormone, and generates a constitutively active construct

[27]. The transcriptional activity of my eight two-domain constructs was measured using a dual luciferase reporter assay.

A luciferase reporter assay measures the activity of a transcription factor through the luminescence of luciferase. In our system, the transcription factor expression vector is co-transfected with a *Gaussia* luciferase vector [58], in which GRE has been cloned in the regulatory region. To take into account the differences in cell density and transfection efficiency, we also co-transfect a control vector that encodes *Cypridina* luciferase [59] not under control of the investigated transcription factor. Both luciferases are engineered to be mostly secreted to the medium and are combined in the dual luciferase assay (as illustrated in Fig. 18).

To measure the transcriptional activity of transcription factors in a reporter gene assay, it is important to ensure that the reporter vector is fully bound by the transcription factor. Otherwise, the activity measured at non-saturating concentrations of transcription factor does not represent the relative activity of each transcription factor as they may have different binding affinities to the response element. To do this, a dosage curve was performed for each translational isoform.

To enable an approximation of the binding affinity of the inactive constructs (D1, D2, and D3) to GRE, a competitive transfection assay is also described in this chapter.

5.3 Materials and methods

Mammalian cell expression vectors for the NTDDBD constructs of GR translational isoforms

Plasmid PJ603-A NTDDBD to express human GR A isoform NTD and DBD two-domain construct in U-2 OS cells was optimized for mammalian cell expression, synthesized by DNA 2.0 (Menlo Park, CA), and inserted into the PJ603 mammalian expression vector under CMV promoter control. Plasmids PJ603-B NTDDBD, -C1 NTDDBD, -C2 NTDDBD, -C3 NTDDBD,

-D1 NTDDDBD, -D2 NTDDDBD and -D3 NTDDDBD were made by inserting the codons for each respective isoform amplified from PJ603 A NTDDDBD into the NheI and XhoI sites of the PJ603 vector.

Luciferase reporter vectors

Plasmid GRE2-Gluc, to express secreted *Gaussia* luciferase under the control of two tandem full-length GREs in the promoteR domain, was made by inserting an oligonucleotide containing two tandem full-length GREs, 5'-aattcAGAACAggaTGTTCTgagatccgtagcAGAACAggaTGTTCTgagatccgtagcg-3', into the EcoRI and BamHI sites of the pGluc-miniTK vector (NEB). Plasmid GRE4half-Gluc, to express secreted *Gaussia* luciferase under control of four tandem half site GREs, was made by inserting an oligonucleotide containing four tandem half-site GREs, 5'-aattcAGAACAggagagatcgtagcAGAACAggaagatccgtagcAGAACAggagagatccgtagcAGAACAggaagatccgtagcg-3', into the EcoRI and BamHI sites of pGluc-miniTK vector. Plasmid pCluc-miniTK2 vector (NEB) to express *Cypridina* Luciferase was utilized as an internal control in the co-transfection to account for cell density differences and transfection efficiency differences in each well.

Mammalian cell culture and transfection

U-2 OS cells (American Type Culture Collection) were maintained in modified McCoy's 5a medium (Corning Cellgro) supplemented with 10% fetal bovine serum and 100 U/mL penicillin and 100 ug/mL streptomycin. To transfect U-2 OS cells at about 80%-90% confluence, X-tremeGENE HP DNA transfection reagent (Roche) was used at 2 µl per 1µg DNA according to the manufacturer's manual.

Western blots

U-2 OS cells were plated on 6 well plates at a density of 5×10^5 cells/well. After 18 hours, 50 ng of GR isoform expression vector with 450 ng Salmon Sperm DNA (Invitrogen) as transfection boosting reagent was transfected into each well with X-tremeGENE HP DNA transfection reagent (Roche) following the manufacturer's protocol. Twenty-four hours post transfection the medium was changed once. After 48 hours, the cells were collected for the assay.

Cells were scraped from each well with PBS, and pelleted by centrifuging at 1500 rpm for 10 minutes in a bench top centrifuge. Fifty μ L of lysis buffer (8 M urea, 20 mM Tris-HCl, 500 mM NaCl, 1 mM Na₂EDTA, 1 mM EGTA, 1% Triton, 2.5 mM sodium pyrophosphate, 1 mM beta-glycerophosphate, 1 mM Na₃VO₄ and 1 μ g/ml leupeptin, pH 7.5) was added into the cell pellet from each well. To reduce the viscosity, cells were passed through a 26G3/8 syringe needle 10 times and then the cell lysate was centrifuged at 14000rpm for 30 minutes. Supernatant was collected and the total protein concentration was measured by Bradford assay (Biorad). Five μ g of total protein was loaded onto each well of a 4–15% Mini-PROTEAN TGX Precast Gel (BioRad), and separated in 1x Tris/Glycine/SDS gel running buffer. Transfer of the protein from SDS page gel to PVDF film was done in transfer buffer (25 mM Tris-HCl, pH 8.3, 192 mM glycine, 20% methanol) under 120V for 15 minutes. After blocking in 5% nonfat milk for 1 hour, the PVDF film with protein transferred on it was incubated with 5000 fold diluted primary antibody for GR (BD Transduction Laboratories, 611226) or 5000 fold diluted antibody for p150glued (BD Transduction Laboratories, 610473), which served as loading control, at 4 °C overnight. The next morning, after washing with PBST buffer (PBS supplemented with 0.1% Tween 20) three times, the PVDF film was incubated in the 10000 fold diluted HRP-linked anti-mouse IgG (GE healthcare, NA931). The detection was done with Amersham ECL Prime Western blotting reagent

(GE healthcare, RPN2232).

Immunostaining

U-2 OS cells were plated on 6 well plates with 15mm round German coverslips. All the culture and transfection procedures are the same as done for the cells treated for Western blot experiments. After 48 hours, cells were rinsed with PBSM (PBS with 2 mM MgCl₂) three times, and fixed with 4% paraformaldehyde in PBSM at room temperature for 10 minutes. Afterwards, each coverslip was rinsed with PBSM three times again, and quenched with 50 mM NH₄Cl in PBSM. Then the slide was placed in PBSTB (PBS with 0.1% Triton X-100, 1%BSA) for 30 minute at room temperature to permeabilize cells and block nonspecific binding. Thereafter, the slide was incubated with primary rabbit antibody for GR (cell signaling, #3660) which is 5000 fold diluted in PBSTB for 1 hour at room temperature. Then the slide was washed three times with PBSM and then incubated with the Alexa Fluor® 488 Goat Anti-Rabbit IgG which is 600 fold diluted in PBSTB for 30 minutes at room temperature in the dark. Next the slide was incubated in PBSM with 0.2 ug/mL DAPI (Invitrogen) and 5 unit/mL Rhodamine Phalloidin (Invitrogen) for 10 minutes at room temperature in the dark to stain the nuclei and F-actin. After that, each slide was washed with PBSM twice and mounted onto a microscope slide with Fluoromount (Sigma), and kept in the dark for drying.

Images were taken with an inverted light microscope (Axiovert 200, JHU Integrated Imaging Center). All images were taken the same day using the same gain, exposure times, and filter configurations (DAPI, FITC, and Texas Red filters of Sedat3 filter set).

Luciferase dosage curve to measure transcriptional activity of GR translational isoforms

U-2 OS cells were plated on 96 well plates at a density of 3×10^4 cells/well as measured by

hemocytometer. After 18~24 hours, when the cells reach about 80-90% confluence, X-tremeGENE HP DNA transfection reagent (Roche) mediated chemical transfection was carried out. For the transcriptional activity dosage curve, 40 ng GRE2-Gluc, 40 ng pCluc-miniTK2 and up to 5 ng GR expression vector were co-transfected into U-2 OS cells on a 96-well plate. After 48 hours, *Gaussia* luciferase activity and *Cypridina* luciferase activity were measured with the BioLux® *Gaussia* Luciferase Assay Kit (NEB) and the BioLux *Cypridina* Luciferase Assay Kit (NEB) respectively on a TriStar² LB 942 Multidetector Microplate Reader (Berthold) according to the manufacturer's protocols. In each experiment, the *Gaussia* luciferase activity (normalized with the *Cypridina* luciferase activity) was measured in triplicate and averaged. Data were fitted to the dose response function as follows.

$$obs = X_{basal} + \frac{X_{\max imum} - X_{basal}}{1 + \left(\frac{EC_{50}}{C_{GR}} \right)^p}$$

Where *Obs* is the observed luciferase expression level under each transfection condition. X_{basal} is the *Gaussia* luciferase expression level without GR regulation. $X_{\max imum}$ is the *Gaussia* luciferase expression level when the GRE in its promoter region is fully bound by GR. C_{GR} is the amount of GR construct vector transfected into U-2 OS cells. EC50 is the GR expression vector transfection amount when the *Gaussia* luciferase expression level is half of its maximum. p: Hill coefficient. In this equation, p is a practical value introduced in the fitting equation, to transform DNA vector amount to protein expression amount, to consider that different isoforms may have different binding cooperativities, different degradation rates, different expression levels, and different nuclear localization efficiencies.

Competitive transfection assay to measure the relative binding affinity

For the competitive transfection assay, 40 ng of GRE4half-Gluc, 40 ng of pCluc-miniTK2, 3 ng of PJ603 C3-NTDDBD, and up to 16 ng of plasmids coding for the competitors were co-transfected into U-2 OS cells on 96-well plates. After 48 hours, the luciferase expression level was measured as described above. The competitive binding data were fitted to the following equation.

$$obs = X_{C3} - \frac{X_{C3} - X_{Competitor}}{1 + \left(\frac{EC_{50}}{C_{competitor}}\right)^p}$$

Where *Obs* is the observed luciferase expression level under each transfection condition. X_{C3} is the expression level of *Gaussia* luciferase under the control of the two-domain constructs of C3 isoform. $X_{Competitor}$ is the expression level of *Gaussia* luciferase under the control of the competitor. $C_{Competitor}$ is the amount of competitor vector transfected into U-2 OS cells. EC_{50} is the competitor vector transfected when the *Gaussia* luciferase expression level is half of the sum of X_{C3} and $X_{Competitor}$. p is the Hill coefficient. In this equation, p is a practical value introduced in the fitting equation, and its purpose is the same as described in the previous paragraph.

5.4 Results and discussion

U-2 OS cells transiently transfected with each isoform can express each isoform by itself

As shown in the Western blot (Fig. 19), U-2 OS cells transiently transfected with the NTDDBD construct of the GR translational isoforms can express each isoform with a similar expression level. In human cells, the Kozak context for the first start codon of GR is not strong, and consequently, multiple translational isoforms are expressed through ribosome leaky and/or shunting mechanisms [25]. In this study, the codon was optimized and for each isoform, the initial start codon was placed in an optimal Kozak context, gacaccAUGg.

Different GR translational isoforms have similar nuclear translocation efficiencies

As shown in the multicolor immunostaining (Fig.20a), the constitutively active constructs of the eight translational isoforms reside both in nuclei (as evidenced by the overlap of green and blue) and in the cytoplasm (as shown by the overlap of green and red). Nuclear percentage is calculated by dividing the intensity of the green dye overlapped with blue dye with the total green dye intensity as shown in Fig.20a. Between four and seven cells were used for each isoform for quantification. Average values and standard errors are reported in the graph (Fig.20b). It can be seen that different translational isoforms have indistinguishable nuclear localization efficiencies.

Different translational isoforms have different transcriptional activities

Each GR vector was separately transfected in increasing amounts to measure the transcriptional activity of each GR translational isoform. The maximum transcriptional activity and EC50 for each isoform were fitted from these dosage curves with the dose response function (listed in Table 4). As shown in Fig. 21, the two-domain constructs of the eight translational isoforms have different activities, and show similar patterns to those reported for the full-length construct under hormone stimulation [25]. A, B, C1, C2 and C3 constructs, whose F domain is complete (as shown in Fig. 1), all have significant transcriptional activities. Their activity is highly sensitive to the variable length of the R domain. Remarkably, the C3 isoform, where the R domain is completely truncated, has more than 3 fold higher activity compared with the other isoforms. D1, D2 and D3 isoforms, in which the activation function 1 (AF1) is fully truncated (as illustrated in Fig. 2), have very low transcriptional activities. As expected, DBD by itself did not have transcriptional activity.

Correlation between the binding affinity measured in vitro and the EC50 fitted from luciferase

dosage curve

From the luciferase assay dosage curve (Fig. 21), the EC50 can be fitted for the A, B, C1, C2 and C3 isoforms (listed in Table 4). Pearson correlation analysis was performed on the EC50 and the binding affinity (Chapter 4, Table 3) for the five active isoforms. As shown in Fig. 22, there is significant correlation between the EC50 and binding affinity. The importance of this correlation is two-fold. First, the EC50 fitted from the luciferase dosage curve approximates the binding of each GR construct to the GRE cloned into the luciferase vector. Second, it demonstrates that the *in vitro* measured binding affinity provides an accurate representation of the *in vivo* behavior, suggesting that in spite of the fact that the protein expressed in mammalian cells can undergo post translational modifications and may have different properties compared to the proteins purified from *E. coli*, these differences do not impact the interpretation of results.

Competitive transfection assays can approximate the binding affinity of the inactive isoforms

As the D1, D2 and D3 isoforms have very low activity, it is not possible to accurately fit their EC50 from the luciferase assay dosage curve (shown in Fig. 21). To approximate the *in vivo* binding affinity of these inactive isoforms, a competitive transfection strategy was designed using the luciferase reporter assay. The amounts of the following components were kept constant in the transfection: GRE controlled *Gaussia* luciferase vector, internal control *Cypridina* Luciferase vector and GR C3 isoform expression vector. The amount of competitor (inactive constructs bearing the DBD) was gradually increased to compete with the C3 isoform for binding to the GRE. A stronger competitor with higher binding affinity will replace the C3 isoform from GRE with a lower dose, thus has a smaller EC50 in the competitive transfection assay. As shown in Fig. 23, the competitive transfection assays were performed on D1, D2, D3 and DBD to compete with C3

isoforms. The fitted EC50 showed significant correlation with the *in vitro* measured binding affinity, as shown in Fig. 24. This suggests that competitive transfection assay can approximate the binding affinity of the inactive isoforms.

Lack of correlation between the transcriptional activity and binding affinity of eight GR translational isoforms

Thus far, it has been shown that the transcriptional activity (listed in Table 4) and binding affinity to GRE (listed in Table 3) can be measured for the two-domain constructs of GR. Comparing the binding affinity to GRE and transcriptional activity did not reveal a significant correlation as shown in Fig. 25. If one examines the binding affinity and transcriptional activity in detail, there seem to be three groups. Remarkably, the C3 isoform has medium binding affinity to GRE and yet it has more than 3 fold greater activity compared with other isoforms. The A, B, C1 and C2 isoforms, which contain the F domain as well as an R domain of different lengths, show medium levels of both transcriptional activity and binding affinity. Minimum activity is shown by the D1, D2 and D3 isoforms, in which the co-regulators' binding sites in AF1 is missing. Surprisingly, two of the D isoforms, D2 and D3 have much higher binding affinity than the other isoforms. Thus they may work as strong competitors to occupy the GR response element (GRE) and regulate the apparent total transcriptional activity of GR in cells. From the binding and activity data, it is clear that GR can exploit the eight translational isoforms to generate a broad spectrum of transcriptional efficiencies.

The lack of correlation between binding affinity to DNA response element and transcriptional activity seems to be a puzzling observation, because binding energetics to the response element is generally believe to determine their specificity and activity [28, 29]. In the

following chapters, the allosteric coupling mechanism will be investigated in an attempt to resolve this puzzling observation.

5.5 Conclusions

The constitutively active constructs of the eight GR translational isoforms had varying activities as measured in the dual luciferase assay dosage curves. The correlation between the EC50 fitted from luciferase assay and the *in vitro* measured binding affinity for the active isoforms confirmed that the *in vitro* measured binding energetics recapitulate the *in vivo* binding. In addition, the competitive transfection strategy described in this chapter provides access to the binding affinity of the inactive isoforms to GRE. The absence of a correlation between the binding affinity to GRE and the transcriptional activity of the eight isoforms suggests that allostery may play an important role in regulating the function of GR translational isoforms.

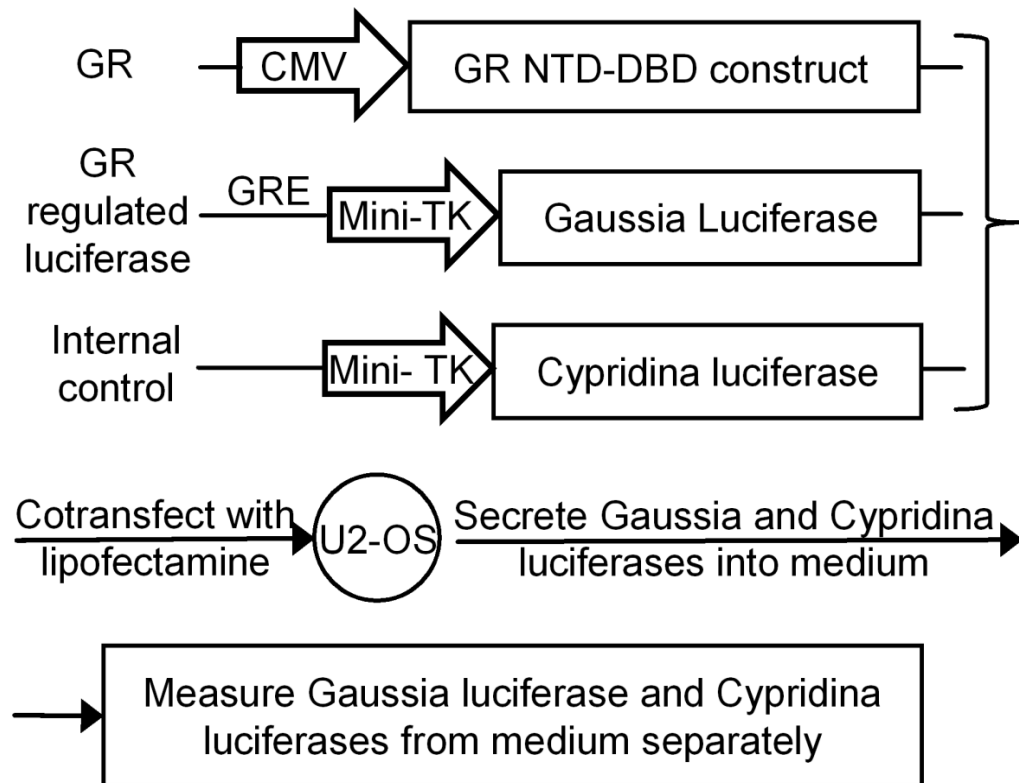


Figure 18 | Co-transfection strategy in the dual luciferase reporter assay.

p150Glued

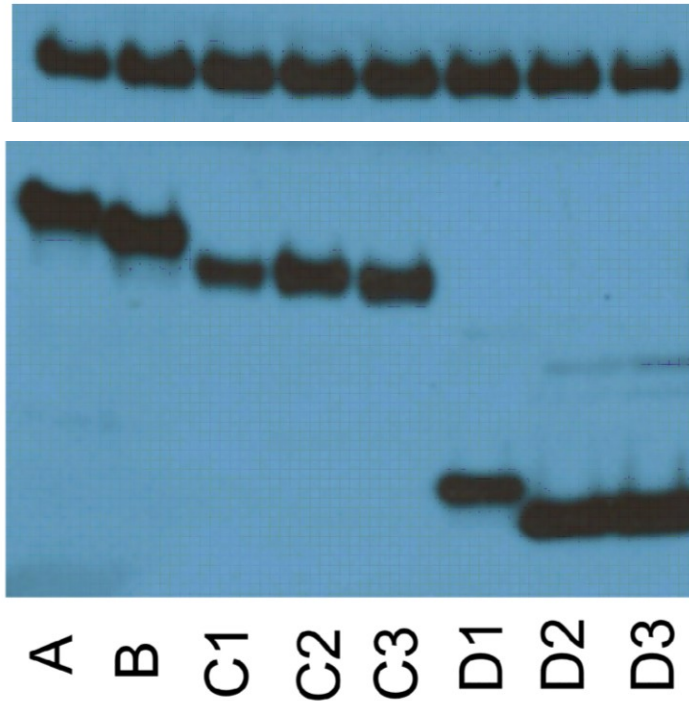


Figure 19 | U-2 OS cells transiently transfected with each isoform can express each single isoform shown by Western blot.

In the upper part of the figure is shown the loading control, p150. In the lower part of the figure is shown the expression level of each GR translational isoform in U-2 OS cells with transient transfection.

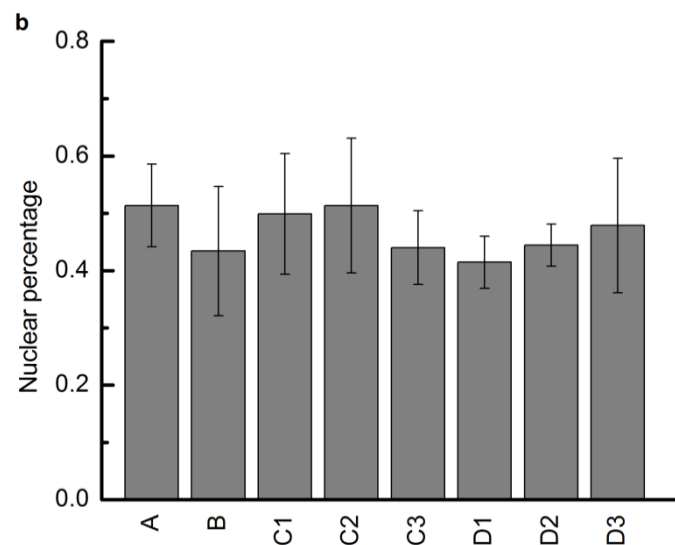
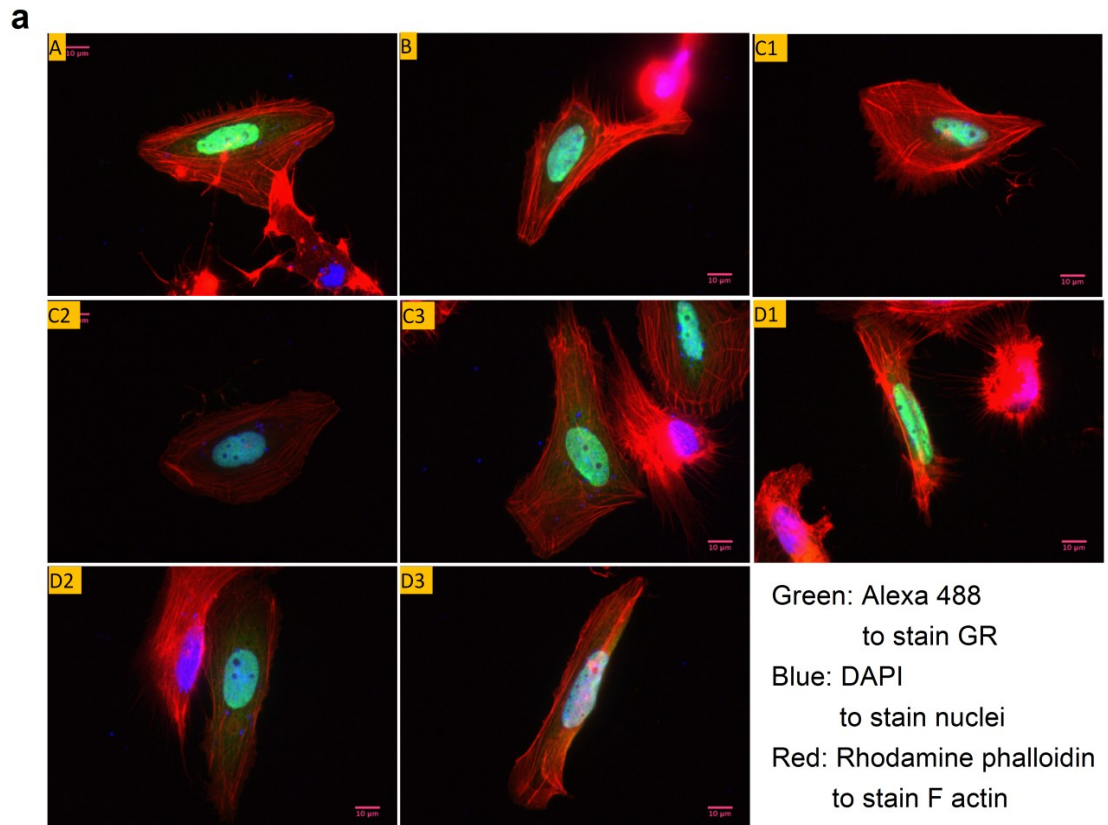


Figure 20 | Immunostaining of U-2 OS cells transfected with GR translational isoforms.

a. Multicolor immunostaining of U-2 OS cells transfected with A, B, C1, C2, C3, D1, D2 and D3 constructs (labeled in top left of each window). Green: Alexa 488 linked goat anti-mouse IgG staining GR. Blue: DAPI staining nuclei. Red: Rhodamine Phalloidin staining F-actin. **b.** Nuclear localization efficiency for the eight GR translational isoforms. Nuclear percentage is calculated by dividing the intensity of the green dye overlapped with blue dye with the total green dye intensity of the cell as shown in *a*. Between four and seven cells were used for each isoform to do the quantification. Average values and standard errors are reported in the graph.

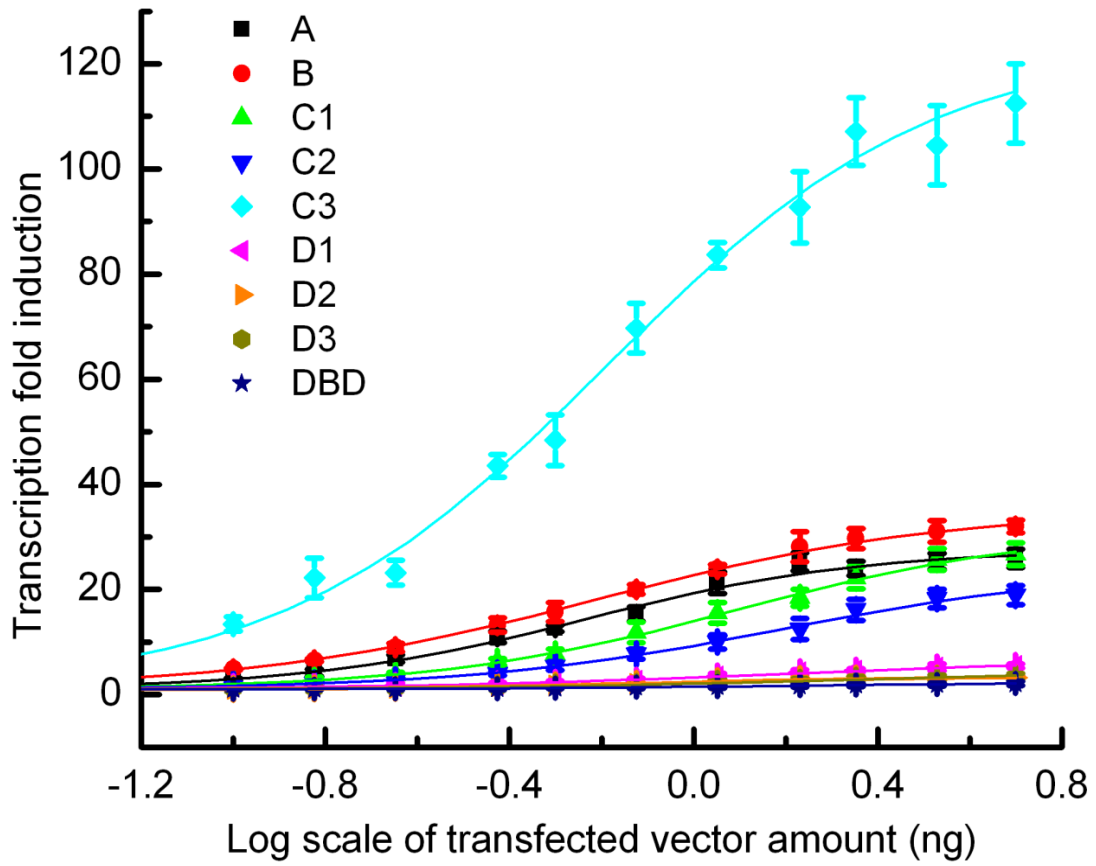


Figure 21 | Dosage curve of the transcriptional activity of the two-domain constitutively active constructs of the eight translational isoforms monitored by dual luciferase reporter assay.

$$obs = X_{basal} + \frac{X_{max} - X_{basal}}{1 + (\frac{EC_{50}}{C_{GR}})^p} \cdot A_{max}$$

Curves fit the data points with the dose response function,

represents the maximum transcriptional activity, and EC50 represents the GR construct amount transfected when the transcriptional activity is half of the maximum. C is the amount of GR construct transfected at each data point. A practical value, p, was introduced in the fitting equation for the following reasons: to transform DNA vector amount to protein expression amount, to account for the probability that different isoforms may have different degradation rates, different expression levels, different nuclear localizations and different binding cooperativities.

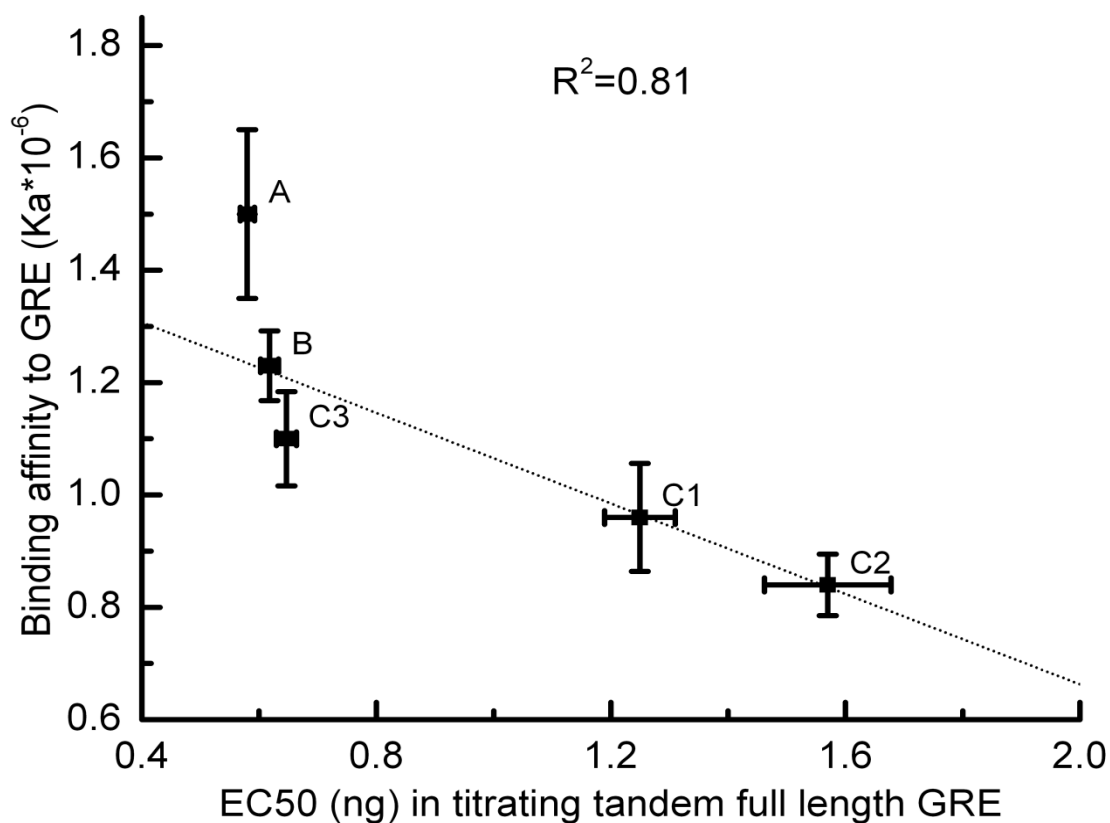


Figure 22 | Correlation between EC50 fitted from *in vivo* dosage curve and *in vitro* measured binding affinity for five active GR translational isoforms.

Correlation between the EC50 fitted from the *in vivo* dosage curve as listed in Table 4 and the *in vitro* binding affinity as shown in table 3, suggesting that the *in vitro* binding affinity represents the *in vivo* binding.

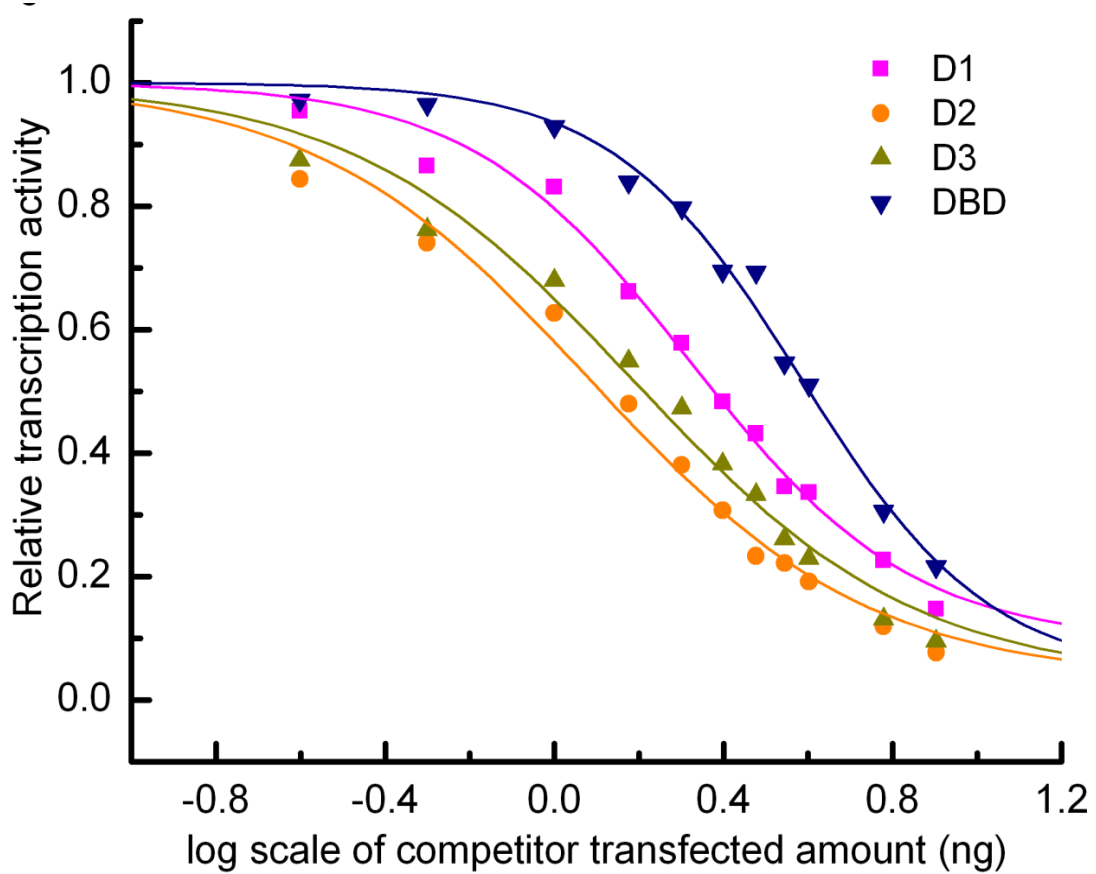


Figure 23 | Competitive binding luciferase assay comparing increasing amounts of D1, D2, D3 and DBD co-transfected with the C3 isoform

The transcriptional activity of C3 isoform in absence of competitors was normalized to 1. Curves

$$F(C) = 1 + \frac{A_{competitor} - 1}{1 + \left(\frac{EC_{50}}{C}\right)^p} \cdot A_{competitor}$$

fit the data points with the dose response function,

represents the transcriptional activity when only 16 ng of competitor is transfected. EC50 represents the competitor construct transfected when the transcriptional activity is half of the C3 maximum. C is the amount of competitor construct transfected at each data point. The value p is a practical value for the same reasons stated in figure 22.

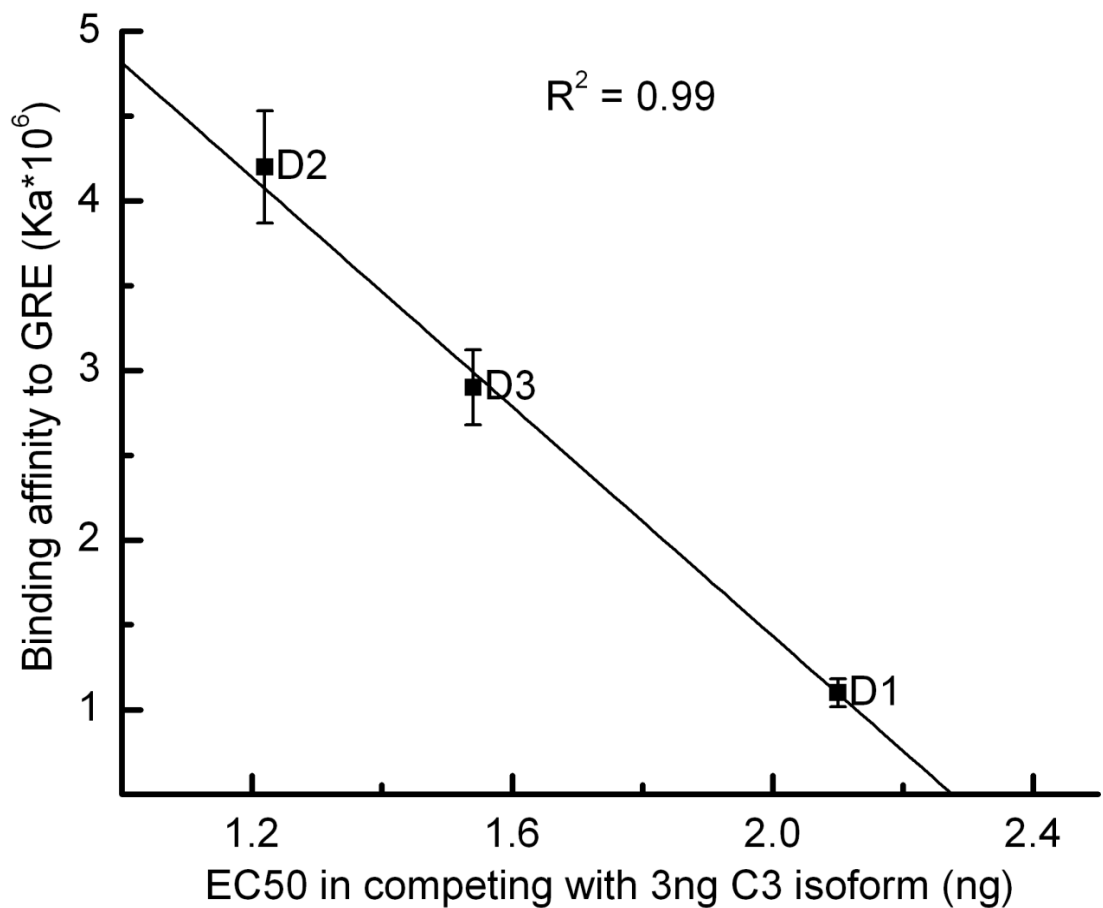


Figure 24 | Correlation between the EC50 fitted from the competitive binding assay and the *in vitro* measured binding affinity for D1, D2 and D3 isoforms.

The correlation demonstrates that the competitive luciferase assay can provide qualitative information about the binding affinity of each construct.

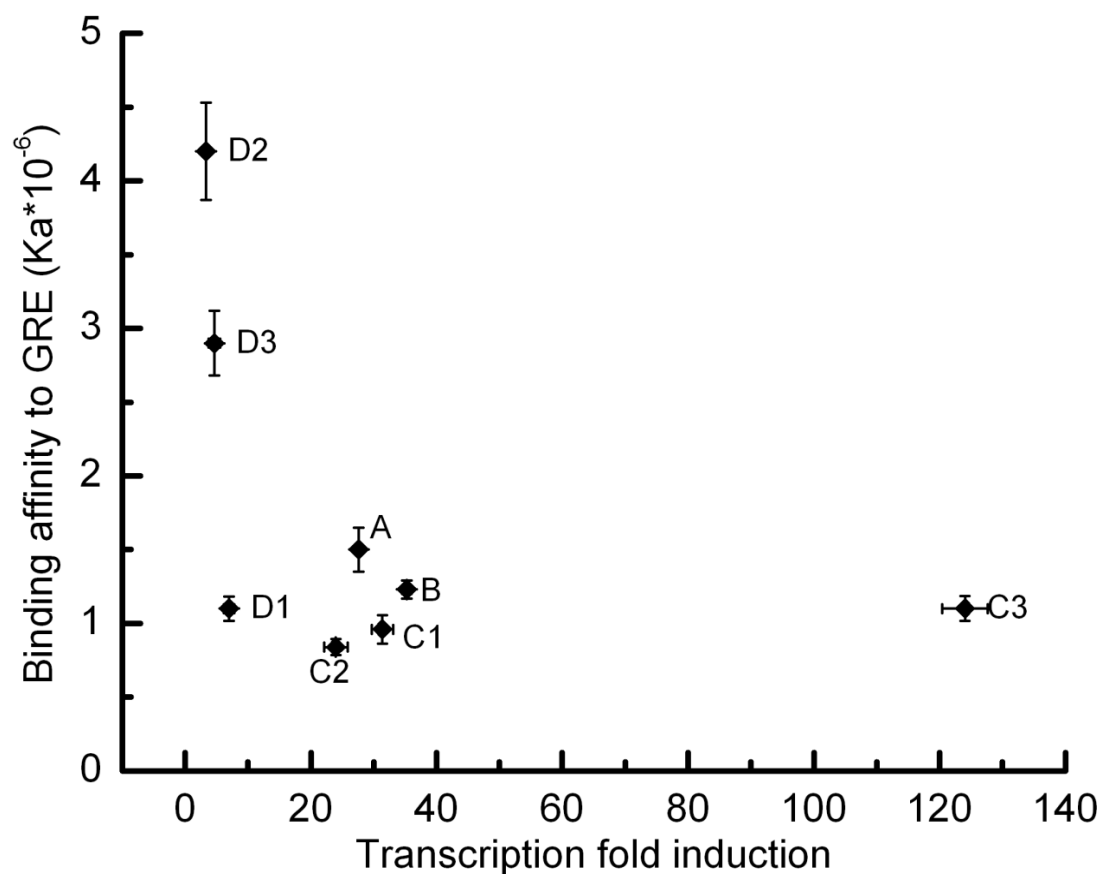


Figure 25 | Lack of correlation between transcriptional activity and binding affinity to GRE of the eight GR translational isoforms.

The transfection activity is fitted from the luciferase assay dosage curve as listed in Table 4, and the binding affinity is measured by monitoring the anisotropy change of the 6FAM labeled GRE as listed in Table 3.

Table 4. Fitted parameters from the luciferase assay dosage curve for A, B, C1, C2 and C3 five active isoforms.

Construct	Transcriptional activity		p		EC50 (ng)	
	Value	Error	Value	Error	Value	error
A	27.64	0.62	1.44	0.05	0.58	0.01
B	35.26	0.79	1.15	0.03	0.62	0.02
C1	31.38	1.70	1.32	0.06	1.25	0.06
C2	23.98	1.87	1.27	0.08	1.57	0.11
C3	124.03	3.64	1.23	0.04	0.65	0.02

Chapter 6-Bidirectional Competing Allosteric Coupling in GR

6.1 Abstract

Previous chapters have demonstrated the unfavorable coupling between the regulatory (R) and functional (F) domains within the NTD, and favorable coupling between the F domain and DBD. Examination of the transcriptional activity and binding affinity data in detail suggests that there is favorable coupling between the R domain and the DBD. To test this hypothesis, R domains of different translational isoforms were directly linked to the DBD through an eleven amino acid flexible linker. The binding affinities of the conjoined constructs were then compared with the linker DBD control construct using the competitive transfection assay. Linking the R domains of the A and B isoforms to DBD significantly increased DBD's binding affinity to GRE, supporting the hypothesis that the R domain is favorably coupled to DBD. As the F and R domains in the NTD are unfavorably coupled to each other, and both are favorably coupled to the DBD, there appears to be a competing allosteric coupling in the constitutively active GR two-domain construct. DBD is favorably coupled to the F domain directly, and unfavorably coupled to the F domain indirectly through its interactions with the R domain. In different translational isoforms, modulating the length of the R domain also regulates the indirect component of the coupling energetics.

6.2 Introduction

In chapter 2, a thermodynamic characterization revealed that the NTD of GR is composed of two functionally distinct regions, a regulatory (R) and a functional (F) domain, which are negatively coupled to each other. In chapter 3, by comparing the TMAO induced folding of C3 NTD (F domain) and C3 NTDDBD (F –DBD), we found that the F domain is positively coupled

to the DBD. We note that these two couplings alone cannot explain why the A isoform (which is composed of R, F and DBD) has higher binding affinity to GRE than the C3 isoform (which is composed of F and DBD). These observations are possible however, if the R domain and DBD are also positively coupled to each other. To test this hypothesis, R domains of the A (GR 1-97), B (GR 27-97), C1 (GR 86-97) and C2 (GR 90-97) isoforms were directly linked to DBD through an eleven amino acid, flexible linker (GTGGSGGSGGS). The binding affinities of these constructs to GRE were compared with DBD through the competitive transfection assay, revealing that the R domain and DBD are indeed favorably coupled. The implications of these results are discussed.

6.3 Materials and methods

Constructs

Amino acids 1-97 of GR were linked to DBD with an eleven amino acid, flexible linker (GTGGSGGSGGS, plasmid PJ603-1-97-11aa-DBD). This plasmid was made by a PCR that deleted the codons for GR 98-420 from PJ603-A NTDDBD and had BamHI and KpnI sticky ends for ligation of an oligonucleotide coding for GTGGSGGSGGS. Plasmid PJ603-27-97-11aa-DBD, -86-97-11aa-DBD, -90-97-11aa-DBD and -11aa-DBD were made by amplifying the appropriate the codons from PJ603-1-97-11aa-DBD and inserting them into the NheI and XhoI site of PJ603 vector. All the single point mutations on GR constructs were made by site directed mutagenesis [60].

Competitive transfection assay

The same as described in Chapter 5.

6.4 Results and discussion

Directly linking R domains of A and B isoforms to DBD significantly increases DBD's binding

affinity to GRE

As shown in Fig.26, linking the R domain of the A or B isoform (R_A or R_B , respectively) to DBD significantly increases the binding affinity of DBD to GRE. The R_A -linker-DBD construct and R_B -linker-DBD constructs have significantly lower EC50 values compared to the linker-DBD construct in the competitive transfection assay. In contrast, linking the R domain of the C1 or C2 isoform to DBD doesn't significantly change the binding affinity of DBD to GRE, as the R_{C1} -linker-DBD and R_{C2} -linker-DBD constructs show similar EC50s to the linker-DBD construct. In summary, the segment 1-85, which is unique to A and B isoforms, is involved in the favorable coupling network between R and DBD. Therefore, changing the length of the R domain in the transcriptionally active isoforms regulates the coupling energetics between the R domain and DBD.

Bidirectional energetic couplings are competitive in constitutively-active, two-domain GR constructs

As summarized in Fig.27, there are two allosteric coupling mechanisms that compete with one another in the two-domain construct of GR. DBD is favorably coupled to the F domain in the NTD through a direct manner, and unfavorably coupled to the F domain via an indirect mechanism that is mediated through interaction with the domain. If the coupling is illustrated in the pie chart shown in Fig.28, we can see that binding of GRE to DBD stabilizes DBD, and via the direct, favorable coupling it will stabilize F domain and favor co-regulator binding to turn on gene transcription. Concomitantly, through an indirect mechanism, stabilization of DBD will stabilize the R domain. However, because of the negative coupling between R and F domains, stabilization of the R domain will consequently destabilize the F domain and disfavor co-regulator binding. In

other words, binding DNA to the same site elicits two different, opposite effect.

6.5 Conclusions

It has been shown that allostery between the R domain and DBD can be probed using a functional assay. Linking the R domains of A and B isoforms to DBD significantly increased DBD's binding affinity to GRE, suggesting that the 1-85 segment of GR, which is unique to GR A and B isoforms, mediates the favorable coupling between R and DBD. Combining this with the coupling data between R and F domains, and between F domain and DBD reveals a complex coupling scheme whereby DNA acting as an allosteric ligand influences transcriptional activity via two competing and opposing effects, which are apparently modulated through the expression of different isoforms. In future chapters this hypothesis is directly challenged.

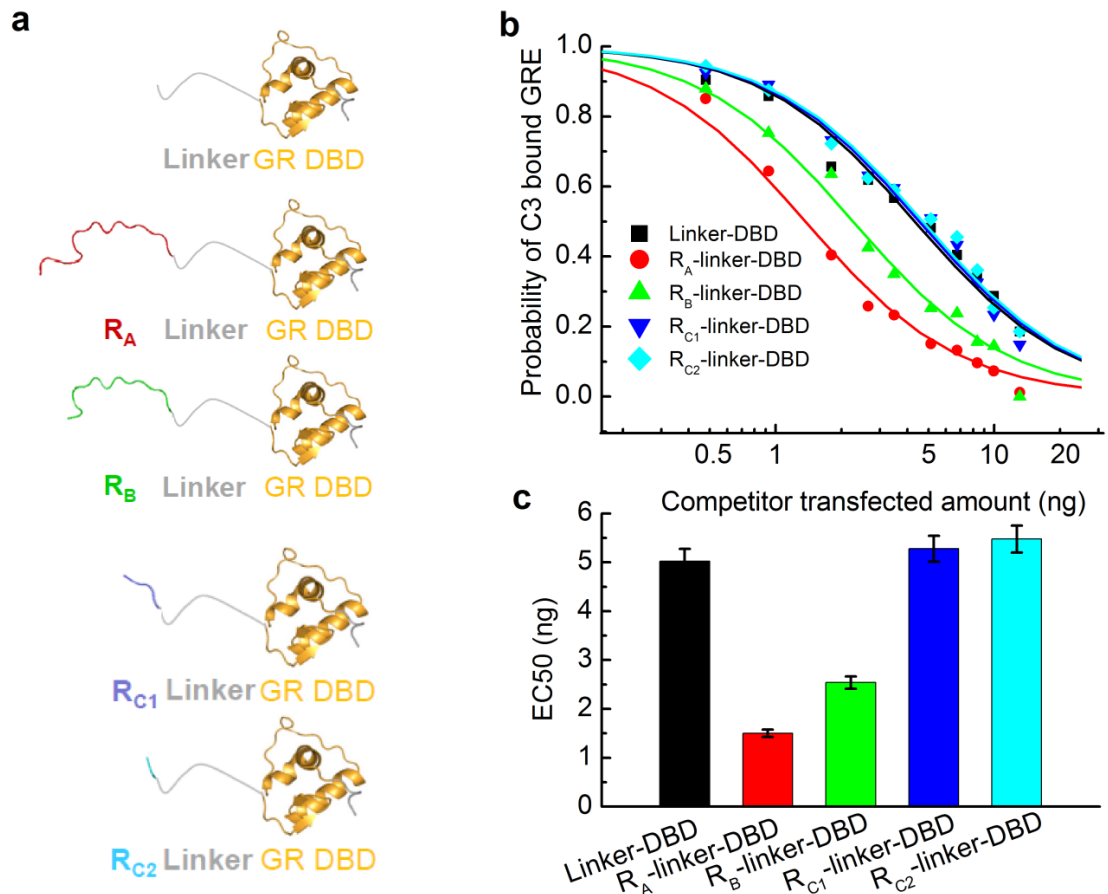


Figure 26 | Competitive transfection assay comparing a construct with the R domain conjoined to the DBD versus the DBD construct.

a. Constructs used to probe the R domain and DBD coupling. R domain of A, B, C1 and C2 isoforms were directly linked to DBD through an eleven amino acid, flexible linker (GGSGGSGGSGT). **b.** Competitive transfection assay curves for the R domain (R_A , R_B , R_{C1} , and R_{C2}) joined to DBD constructs versus the DBD construct. **c.** Fitted EC50 values for each construct from the competitive curves shown in b. As shown linking R domain of the A isoform (GR 1-97) or the B isoform (GR 27-97) to the DBD through a flexible linker (GGSGGSGGSGT) significantly increases the binding affinity of DBD to GRE as shown by the significantly reduced EC50 value. Linking the R domain of the C1 isoform (GR 86-97) or the C2 isoform (GR 90-97) to DBD doesn't significantly influence the binding affinity to GRE. This suggests that the R domain of A and B isoforms are favorably coupled to the DBD.

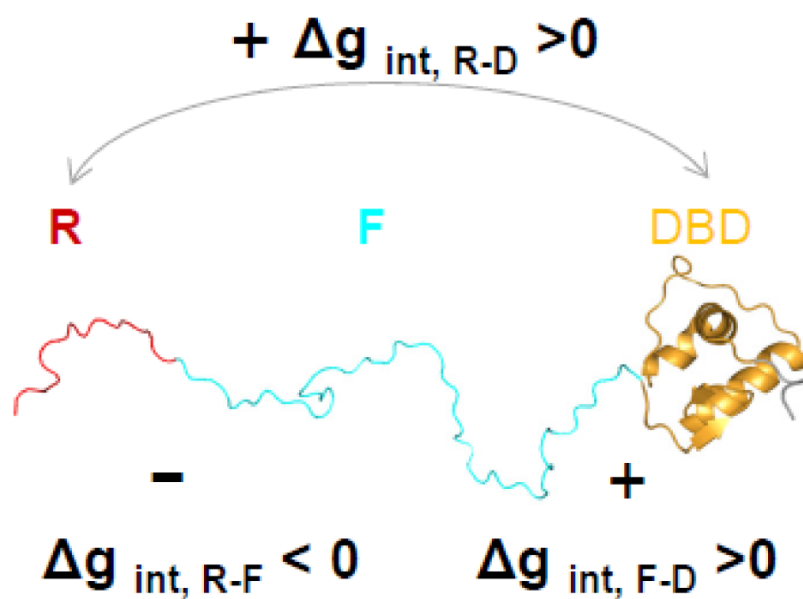


Figure 27 | Summary of the allosteric coupling network in the two-domain construct of GR. Cartoon summary of the thermodynamic coupling scheme in GR two-domain construct, R- and F-domains within the intrinsically disordered NTD are unfavorably coupled to each other, and both of them are favorably coupled to the DBD.

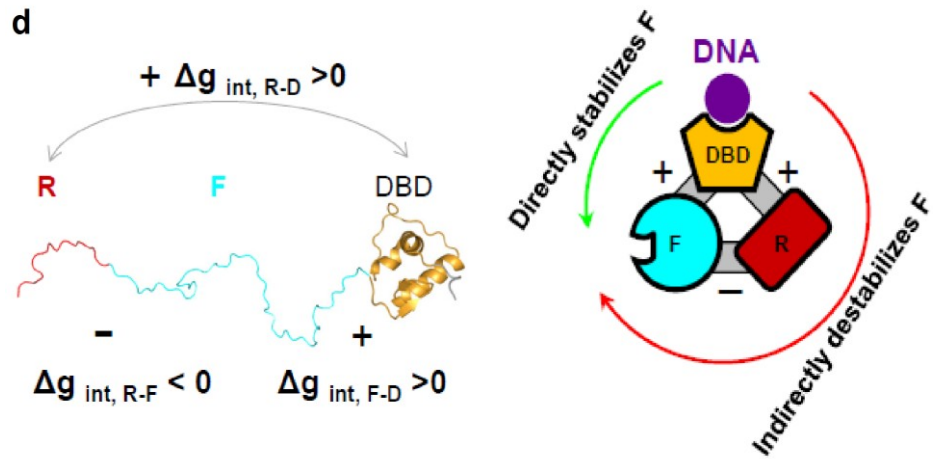


Figure 28 | Bi-directional competing allosteric coupling network when allosteric effector GRE binds to the DNA binding domain

When GRE binds to the DBD, it will stabilize DBD and directly stabilize the F domain. Meanwhile, stabilization of DBD will stabilize the R domain, and consequently destabilize the F domain indirectly.

Chapter 7-Simulation with Ensemble Allosteric Model for GR

Translational Isoforms

7.1 Abstract

In this chapter, an Ensemble Allosteric Model (EAM) is constructed for each GR translational isoform. The biophysical characterization and functional assays provide realistic starting values for each thermodynamic parameter in the model. Based on these inputs, we carried out an exhaustive search of parameter space to satisfy the experimental constraints on activity and binding affinity of the eight isoforms. It was found that with the best fit value of each parameter from the EAM, both simulated binding affinity and transcriptional activity, showed significant correlation with the experimental data. As such, the EAM can be used to gain insight into the observation that binding affinity and transcriptional activity are not correlated for GR translational isoforms.

7.2 Introduction

Previous chapters discussed the biophysical characterization and cell based functional assays that illuminated the allosteric coupling mechanism of GR, as summarized in Fig.27. DBD is favorably coupled to the F domain in the NTD through a direct mechanism, and unfavorably coupled to the F domain via an indirect mechanism that utilizes the R domain. Binding of GRE to DBD stabilizes the DBD, and this will directly stabilize the F domain and favor the binding of co-regulators to turn on gene transcription. Concomitantly, through an indirect manner, stabilization of the DBD will stabilize the R domain, and stabilization of the R domain will destabilize the F domain and disfavor the binding of co-regulators.

The allosteric coupling mechanism discussed above enables us to build an Ensemble

Allosteric Model (EAM) [6, 7] for each translational isoform. A parameter space search was carried out using an experimentally guided range for each parameter.

7.3 Methods

EAM for each two-domain construct of GR translational isoforms

The ensemble allosteric model was built for each two-domain construct of GR. The A, B, C1 and C2 isoforms each contain three domains, consisting of the F domain, the DBD and a variable sized R domain. The C3, D1, D2 and D3 isoforms, on the other hand, contain two domains, which include the DBD and variable sized F domains. In each ensemble, the F domain and R domain can either be in the folded or unfolded conformation, while the DBD can adopt a high affinity or a low affinity state for GRE. With the fully folded state as the reference state, the free energy of the other states can be expressed as the free energy to unfold each region, as well as the free energy to break the interaction between the other folded domains. For instance, the models constructed for the full length A isoform (composed of R domain, F domain and DBD) and the most active C3 isoform (composed of the F domain and DBD) are shown in Fig.29. Based on the free energy, ΔG , of each state, the statistical weight, SW, was calculated using the following equation.

$$SW = \text{Exp} \left(-\frac{\Delta G}{RT} \right)$$

Sum of the statistical weights of each state results in the partition function, Q, for the system.

$$Q = \sum_{i=1}^n SW_i$$

Consequently, the probability of each state, P_i , can be calculated by dividing each state's statistical weight with the partition function Q for the ensemble.

$$P_i = \frac{SW_i}{Q}$$

In the EAM, the probability of states with the F domain folded and DBD in the high affinity conformation represents the relative transcriptional activity for each isoform. The probability of states with DBD in the high affinity conformation represents the relative binding affinity for each isoform. The temperature used in the simulation is 37°C.

Parameter space search with EAM to satisfy experimental constraints

The biophysical measurements and the allosteric coupling mechanism interpreted from experimental results suggested a practical range for each thermodynamic parameter in the model, as shown in Table 5. With a step size of 0.5 kcal/mol, combinations of the free energy terms were exhaustively explored and the probabilities for the functional states and the probabilities for the binding competent states in each isoform's ensemble were determined for each parameter combination. If the calculated results satisfied the constraints based on experimental observations (as shown in Table 6), the corresponding parameter set was defined as a positive hit and the parameters voted. For detailed derivation please see appendix 2.

7.4 Results and discussion

The competing energetic coupling is crucial to satisfy all experimental constraints

The box chart of the stability of each region is shown in Fig. 30. Also shown are the following: the positive coupling energy between the F domain and DBD, the negative coupling energy between the R and F domain, and the positive coupling energy between R domain and DBD from simulation. This indicates that the coupling energetics are crucial to satisfy all the experimental binding affinity and transcriptional activity constraints. In addition, it further supports the validity of the coupling model we proposed from the experimental results.

EAM can capture the absence of a correlation between the transcriptional activity and binding

affinity for GR translational isoforms

With the best fit value of each parameter from simulation (as shown in Fig.30), the simulated transcriptional activity and binding affinity can be calculated for each isoform with the full length A isoform as a reference. As shown in Fig.31, both the simulated transcriptional activity and binding affinity have significant correlation with experimentally measured values. Thus the absence of a correlation between the binding affinity and transcriptional activity of the GR translational isoforms can be understood with EAM.

7.5 Conclusions

In this chapter, EAM was built for each two-domain construct of GR. A parameter space search was carried out to determine the allowable range for the free energy of the R domain, F domain, the DBD, and the interaction energy between them to satisfy the experimental constraints. Examining the range of values that satisfy each interaction energy term reveals that the coupling scheme embedded in the EAM successfully captures the experimental constraints and faithfully reproduces both the transcriptional activity and binding affinity data. The ability of the EAM to capture the bi-directional cooperativity suggests that the model can be used to explore in detail the thermodynamic mechanism of allosteric control utilized by IDPs.


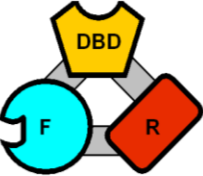
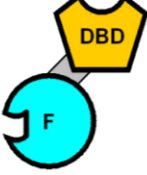

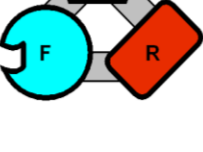
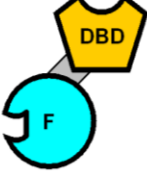

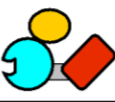

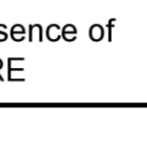


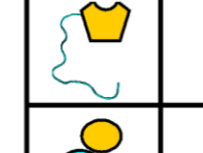
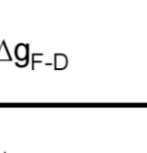
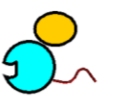

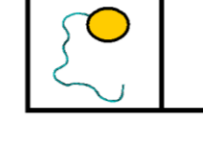
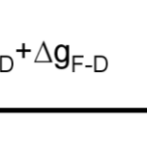





A	ΔG in absence of GRE	A NTDDBD		C3 NTDDBD	
	0				
	$\Delta G_F + \Delta g_{F-D} + \Delta g_{F-R}$				
	$\Delta G_R + \Delta g_{R-D} + \Delta g_{F-R}$				
	$\Delta G_D + \Delta g_{R-D} + \Delta g_{F-D}$				
	$\Delta G_F + \Delta G_R + \Delta g_{R-D} + \Delta g_{F-D} + \Delta g_{F-R}$				
	$\Delta G_F + \Delta G_D + \Delta g_{R-D} + \Delta g_{F-D} + \Delta g_{F-R}$				
	$\Delta G_R + \Delta G_D + \Delta g_{R-D} + \Delta g_{F-D} + \Delta g_{F-R}$				
	$\Delta G_F + \Delta G_R + \Delta G_D + \Delta g_{R-D} + \Delta g_{F-D} + \Delta g_{F-R}$				
					
		C3		ΔG in absence of GRE	
				0	
				$\Delta G_F + \Delta g_{F-D}$	
				$\Delta G_D + \Delta g_{F-D}$	
				$\Delta G_F + \Delta G_D + \Delta g_{F-D}$	

Figure 29 | Ensemble allosteric model for A and C3 isoforms

Table 5. Practical parameter range used in parameter space search

	Lower bound (kcal/mol)	Upper bound (kcal/mol)
ΔG_{DBD}	-2	2
$\Delta G(\text{R}_\text{A})$	-2.5	2.5
$\Delta G(\text{R}_\text{B})$	-2.5	2.5
$\Delta G(\text{R}_{\text{C1}})$	-2.5	2.5
$\Delta G(\text{R}_{\text{C2}})$	-2.5	2.5
$\Delta G(\text{F})$	-7	-7
$\Delta g_{\text{int}}(\text{F}, \text{DBD})$	5	7
$\Delta g_{\text{int}}(\text{R}_\text{A}, \text{DBD})$	0	2
$\Delta g_{\text{int}}(\text{R}_\text{B}, \text{DBD})$	0	2
$\Delta g_{\text{int}}(\text{R}_{\text{C1}}, \text{DBD})$	0	0
$\Delta g_{\text{int}}(\text{R}_{\text{C2}}, \text{DBD})$	0	0
$\Delta g_{\text{int}}(\text{R}_\text{A}, \text{F})$	-3	0
$\Delta g_{\text{int}}(\text{R}_\text{B}, \text{F})$	-3	0
$\Delta g_{\text{int}}(\text{R}_{\text{C1}}, \text{F})$	-3	0
$\Delta g_{\text{int}}(\text{R}_{\text{C2}}, \text{F})$	-3	0

Table 6. Experimental constraints for relative activity and binding affinity for each isoform with full length A isoform as refernece in the parameter space search

Ka Activity	A	B	C1	C2	C3
A		1~1.5	1~2.5	1~2.5	1~2
B	1~1.5		1~2.5	1~2.5	1~2
C1	1~1.5	0.67~1		1~2.5	0.5~1
C2	0.5~1	0.5~1	0.5~1		0.5~1
C3	3~5	2.5~5	2.5~5	3~6.6	

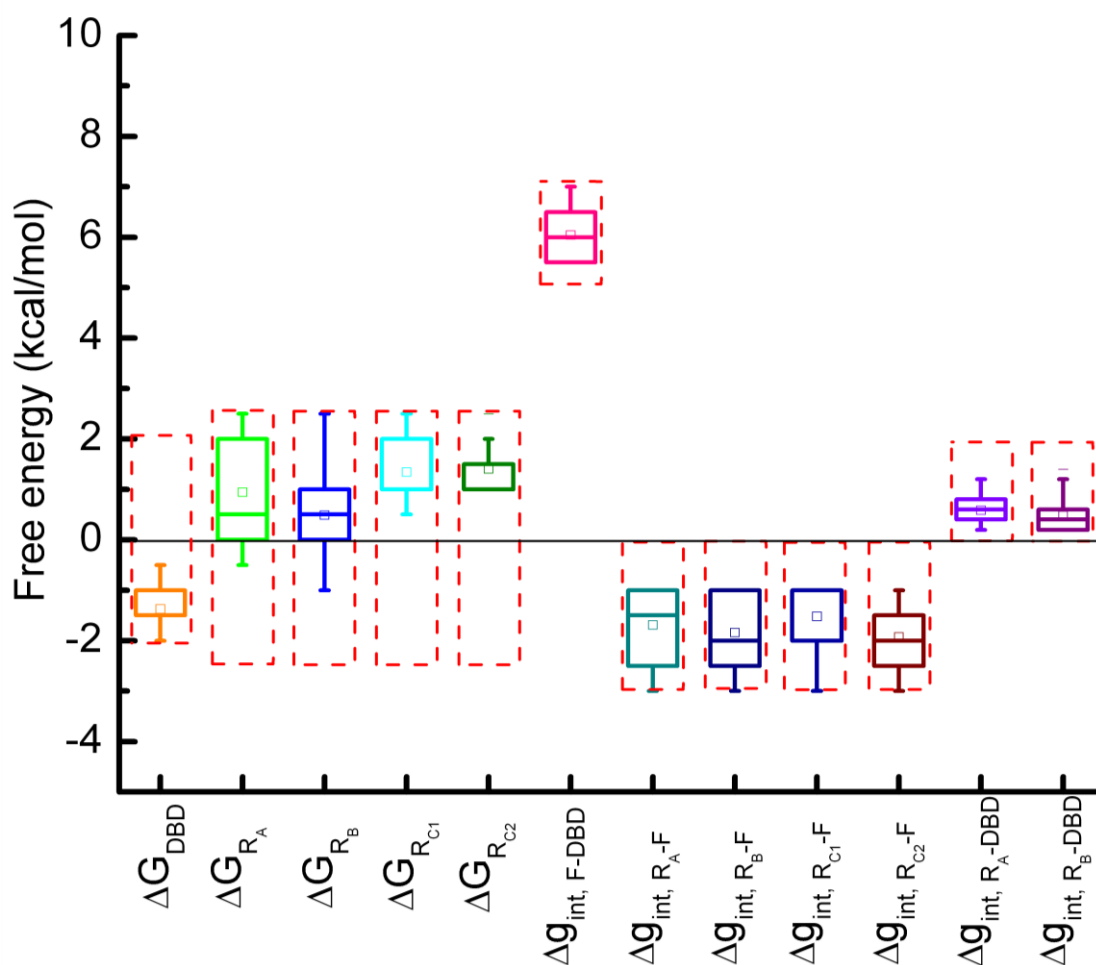


Figure 30 | Box chart of satisfied parameter range given experimental constraints on the Ensemble Allosteric Model.

Notice that the competing coupling energetics are crucial to satisfy all experimental constraints. In each box, the maximum, the 75 percentile, the median, the 25 percentile, and the minimum are shown as lines. The average values are shown as squares. The red dashed lines denote the input parameter range in the parameter space search.

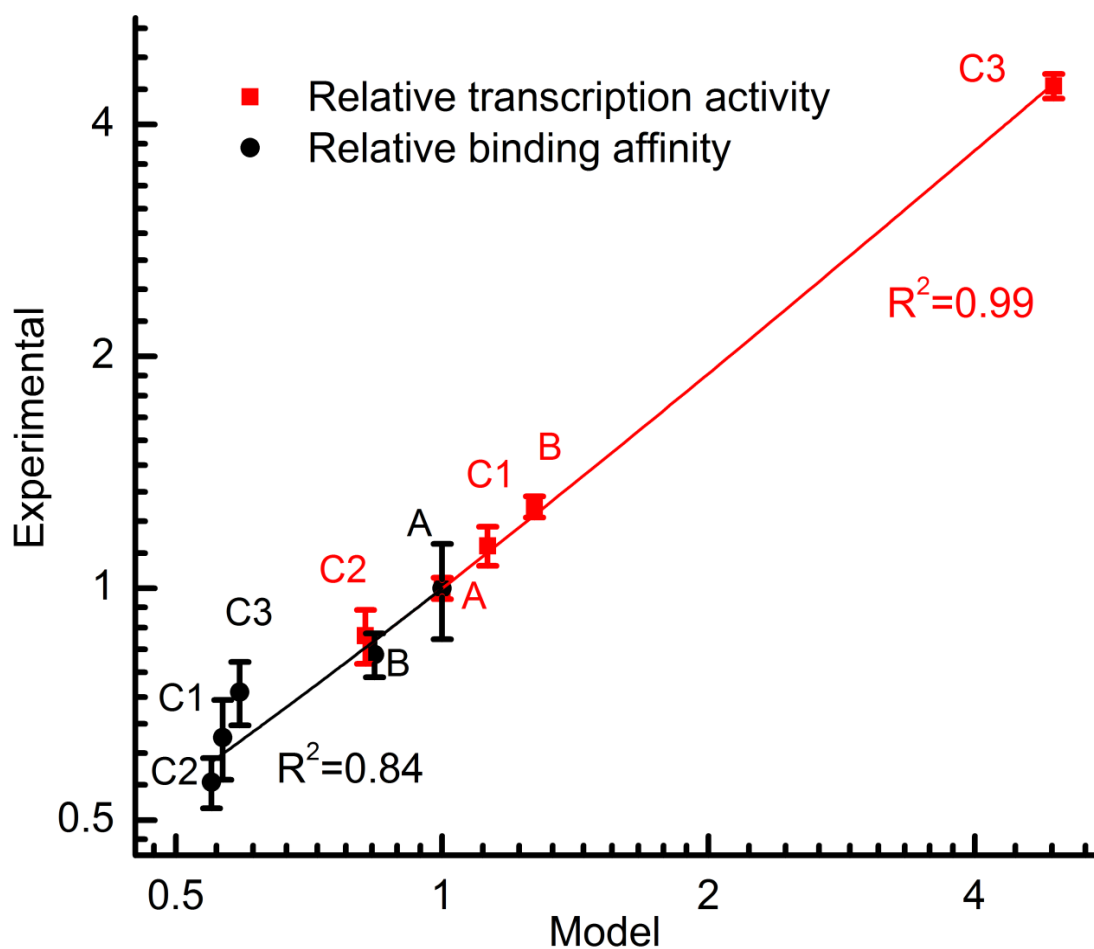


Figure 31 | Correlation between the simulated relative transcriptional activity and binding affinity with experimentally measured values.

Simulated values for transcriptional activity and binding affinity of each isoform are based on the ensemble averaged contribution of all the states in the ensemble. The correlation coefficient between the simulated data and experimental data is the Pearson correlation.

Chapter 8- Applications of EAM: Predictions of Mutational Effects

8.1 Abstract

In this chapter, the ability of the EAM to predict mutational effects is explored. From a thermodynamic perspective, introducing a mutation to a domain allosterically coupled with other domains may either, (1) change the stability of the domain, or (2) weaken the coupling energetics between domains. In both of these two scenarios, influences on transcriptional activity and binding affinity can be predicted from the EAM. These predictions make it possible and straightforward to interpret the mutagenesis effect when pinpointing the molecular basis of allosteric coupling.

8.2 Introduction

Identifying the residues involved in the allosteric coupling network is important for allosteric drug design, as well as obtaining a fundamental understanding of allostery. Computationally, the molecular basis of coupling is sometimes predicted by statistical coupling analysis using a large set of sequence alignments[61]. This is hardly applicable to ID regions as there is usually low sequence conservation of the ID regions within a protein family [62]. Experimentally, the molecular basis of allostery is usually studied by examining the effects of point mutations on thermodynamic binding or function [63]. In a complicated allosteric protein with multiple, coupled domains it is nontrivial to interpret mutagenesis results. However, expected mutational scenarios can be predicted from the EAM of various GR isoforms, which makes pinpointing the molecular basis of allosteric coupling through mutagenesis possible and straightforward.

Introducing a mutation at one region of a protein may cause different effects: a protein

solubility change, a cofactor binding affinity change, a protein thermodynamic stability change, as well as a change in the coupling energetics between regions. Among these, protein solubility changes can be gauged by Western blot, and the likelihood of a change in binding affinity for a binding partner can be largely minimized by avoiding mutations in the annotated binding sites. Our EAM model can predict the effects of mutations perturbing inter-domain allostery or altering thermodynamic stability. These predictions provide us with anticipated results when screening for the molecular basis of the coupling between the R domain and F domain, F domain and DBD, as well as the R domain and DBD, as discussed in the future chapters.

8.3 Methods

Perturbation of GR EAM with mutagenesis introduced to R domain, F domain or DBD

In the EAM, the probability of the states with both F domain in the folded conformation and DBD in the high affinity state represents the transcriptional activity. The probability of the states with the DBD in the high affinity state represents the binding affinity. Thus, both activity and binding affinity are functions of the thermodynamic parameters in the system, which can be expressed as follows:

$$Activity = F1(\Delta G_R, \Delta G_F, \Delta G_{DBD}, \Delta g_{int, R-F}, \Delta g_{int, F-DBD}, \Delta g_{int, R-DBD})$$

$$Affinity = F2(\Delta G_R, \Delta G_F, \Delta G_{DBD}, \Delta g_{int, R-F}, \Delta g_{int, F-DBD}, \Delta g_{int, R-DBD})$$

To simulate the scenario that a mutation influences the stability of a target region, a $\Delta\Delta G$ term is introduced to the respective region's stability term. To simulate the scenario that a mutation influences the allosteric coupling network between two regions, a $\Delta\Delta g_{int}$ term is introduced to the interaction energy term of the two coupled regions. With a mutation introduced to the R domain as an example, its thermodynamic effect may be one of the three cases: (I) changing R domain

stability; (II) influencing the R domain and DBD coupling; (III) influencing R domain and F domain coupling. In case I, the influence on transcriptional activity with wild type (WT) activity as reference can be expressed as:

$$\text{Activity (Mutant/WT)} = F1(\Delta 1_R + \Delta \Delta G_R, \Delta G_F, \Delta G_{DBD}, \Delta g_{int, R-F}, \Delta g_{int, F-DBD}, \Delta g_{int, R-DBD}) / F1(\Delta 1_R, \Delta G_F, \Delta G_{DBD}, \Delta g_{int, R-F}, \Delta g_{int, F-DBD}, \Delta g_{int, R-DBD})$$

The influence on binding affinity to GRE with wild type (WT) binding affinity as reference can be expressed as:

$$\text{Affinity(Mutant/WT)} = F2(\Delta 2_R + \Delta \Delta G_R, \Delta G_F, \Delta G_{DBD}, \Delta g_{int, R-F}, \Delta g_{int, F-DBD}, \Delta g_{int, R-DBD}) / F2(\Delta 2_R, \Delta G_F, \Delta G_{DBD}, \Delta g_{int, R-F}, \Delta g_{int, F-DBD}, \Delta g_{int, R-DBD})$$

Similarly, in case II, the influence on transcriptional activity can be expressed as:

$$\text{Activity (Mutant/WT)} = F1(\Delta 1_R, \Delta G_F, \Delta G_{DBD}, \Delta g_{int, R-F}, \Delta g_{int, F-DBD}, \Delta g_{int, R-DBD} + \Delta \Delta g_{int, R-DBD}) / F1(\Delta 1_R, \Delta G_F, \Delta G_{DBD}, \Delta g_{int, R-F}, \Delta g_{int, F-DBD}, \Delta g_{int, R-DBD})$$

And the influence on binding affinity can be expressed as

$$\text{Affinity (Mutant/WT)} = F2(\Delta 2_R, \Delta G_F, \Delta G_{DBD}, \Delta g_{int, R-F}, \Delta g_{int, F-DBD}, \Delta g_{int, R-DBD} + \Delta \Delta g_{int, R-DBD}) / F2(\Delta 2_R, \Delta G_F, \Delta G_{DBD}, \Delta g_{int, R-F}, \Delta g_{int, F-DBD}, \Delta g_{int, R-DBD})$$

Likewise, in case III, the influence on transcriptional activity can be expressed as:

$$\text{Activity (Mutant/WT)} = F1(\Delta 1_R, \Delta G_F, \Delta G_{DBD}, \Delta g_{int, R-F} + \Delta \Delta g_{int, R-F}, \Delta g_{int, F-DBD}, \Delta g_{int, R-DBD}) / F1(\Delta 1_R, \Delta G_F, \Delta G_{DBD}, \Delta g_{int, R-F}, \Delta g_{int, F-DBD}, \Delta g_{int, R-DBD})$$

And the influence on binding affinity can be expressed as

$$\text{Affinity(Mutant/WT)} = F2(\Delta 2_R, \Delta G_F, \Delta G_{DBD}, \Delta g_{int, R-F} + \Delta \Delta g_{int, R-F}, \Delta g_{int, F-DBD}, \Delta g_{int, R-DBD}) / F2(\Delta 2_R, \Delta G_F, \Delta G_{DBD}, \Delta g_{int, R-F}, \Delta g_{int, F-DBD}, \Delta g_{int, R-DBD})$$

8.4 Results and discussion

Perturbations of GR EAM with mutagenesis introduced to R domain of different translational isoforms

A, B, C1 and C2 isoforms have R domains of different lengths. As shown in Fig. 32a, if a

mutation influences the R domain stability, it will significantly influence the transcriptional activity of all four isoforms, but not have a significant effect on binding affinity (shown in Fig. 32b). If, on the other hand, a mutation influences R domain and DBD coupling, it will influence the transcriptional activity and binding affinity on A and B isoforms only modestly (shown in Fig. 32c&d). If a mutation on C2 influences R and F domain coupling, it will influence the transcriptional activity significantly while affecting the binding affinity modestly (shown in Fig. 32e & f). The same mutation on the other hand will not have a large influence on the activity or the binding affinity of A, B or C1 isoforms (shown in Fig. 32 e & f).

As these mutational scenarios suggest a quantitative model of allostery for the different isoforms, it provides a tool from which the residues critical for coupling can be identified based on their mutational effect.

Perturbations of GR EAM with mutagenesis introduced to F domain of different translational isoforms

All five active isoforms A, B, C1, C2 and C3 have the same F domain. However, in the context of different R domains in each isoform they each respond differentially to perturbations within the F domain. As shown in Fig33.a&b, a mutation destabilizing the F domain will decrease transcriptional activity of all five isoforms, and will moderately decrease the binding affinity to GRE. Likewise, a mutation within the F domain that weakens F domain and DBD coupling will also decrease transcriptional activity of all five isoforms, and modestly decrease binding affinity (shown in Fig. 33c&d). In addition, mutations within the F domain influencing R and F domain coupling have a much more pronounced effect on the transcriptional activity of C2 isoform than on the activity of the other isoforms (shown in Fig. 33e&f). This indicates that by measuring

transcriptional activity change in the A, B, C1 and C2 isoforms upon mutagenesis within the F domain will provide a means of resolving which residues are involved in mediating coupling between F and R domains.

Mutagenesis within the F domain on different isoforms will not distinguish between the different scenarios influencing F domain stability and perturbing F domain and DBD coupling. The two scenarios have similar effects on both transcriptional activity and binding affinity to GRE. Thus to pinpoint the residues within F domain that are allosterically coupled to the DBD, it is necessary to employ a control construct. For example, a control in which the F domain is tethered to the DBD of another transcription factor provides a gauge of the effect of the mutation on F domain stability.

Perturbations of GR EAM with mutagenesis introduced to DBD of different translational isoforms

All the five active isoforms, A, B, C1, C2 and C3, have the same DBD. As predicted from EAM, mutations influencing DBD stability (the free energy difference between the high affinity and low affinity state to DNA) will have little influence on the transcriptional activity (shown in Fig. 34a); however, it will strongly influence the binding affinity to GRE (shown in Fig. 34b). In contrast, mutations weakening F and DBD coupling will strongly influence the transcriptional activity (shown in Fig. 34c), and moderately regulate the binding affinity to DNA (shown in Fig. 34d). Mutations within the DBD may also influence the R domain and DBD allosteric coupling. In this case, the mutations will only moderately influence both transcriptional activity and binding affinity of A and B isoforms (shown in Fig. 34e&f). As such, it is possible to distinguish these three scenarios by measuring both transcriptional activity and binding affinity changes upon

mutation.

Thus, measuring the DBD's mutational effect on both transcriptional activity and binding affinity on the C3 isoform, which is only composed of F domain and DBD, will help to pinpoint the residues involved in the F domain and DBD coupling network.

8.5 Conclusions

In this chapter, thermodynamic perturbation of GR EAM was carried out and the effects on transcriptional activity and binding affinity were simulated. These predictions are useful for distinguishing different scenarios when a mutation is introduced to a system, which makes pinpointing the molecular basis of allostery in complicated systems possible and straightforward.

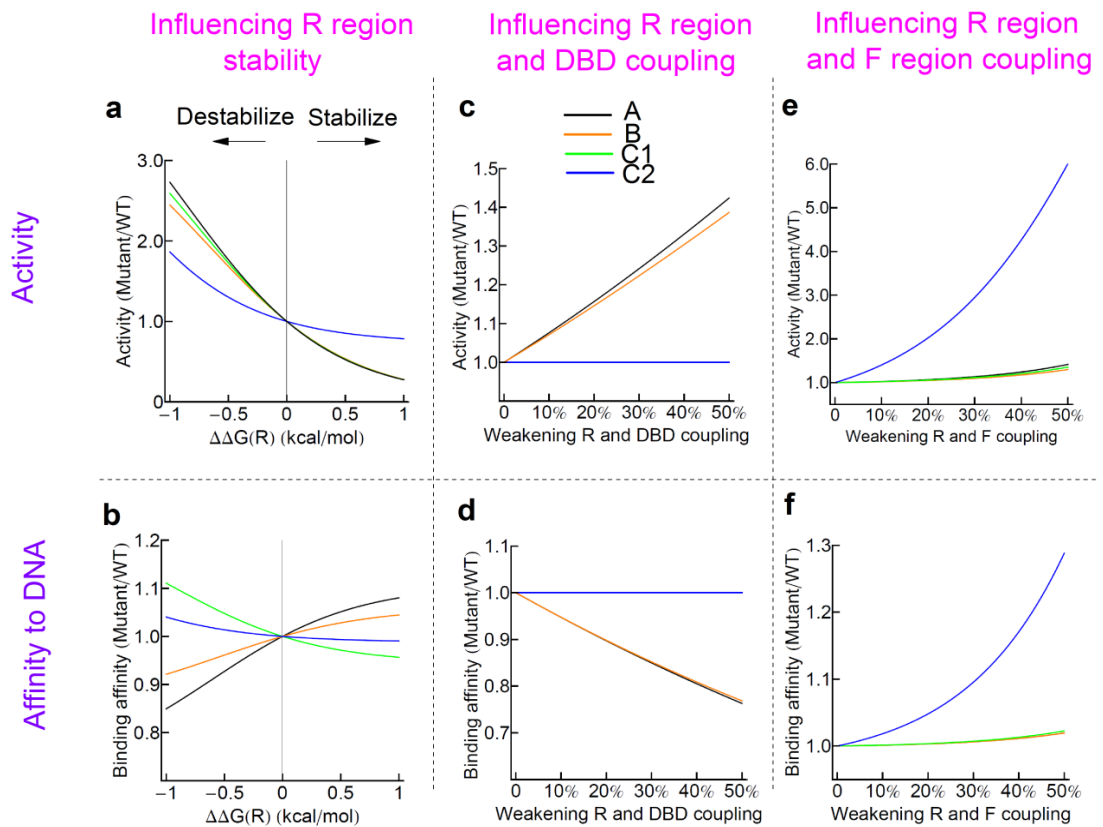


Figure 32 | Predicted influence on transcriptional activity and binding affinity to GRE when mutations are introduced to R domain in different isoforms.

Three possible thermodynamic scenarios may happen when a mutation is introduced into the R domain: influencing R domain stability (shown in a, b), influencing R domain and DBD coupling (shown in c, d), and influencing R and F coupling (shown in e, f).

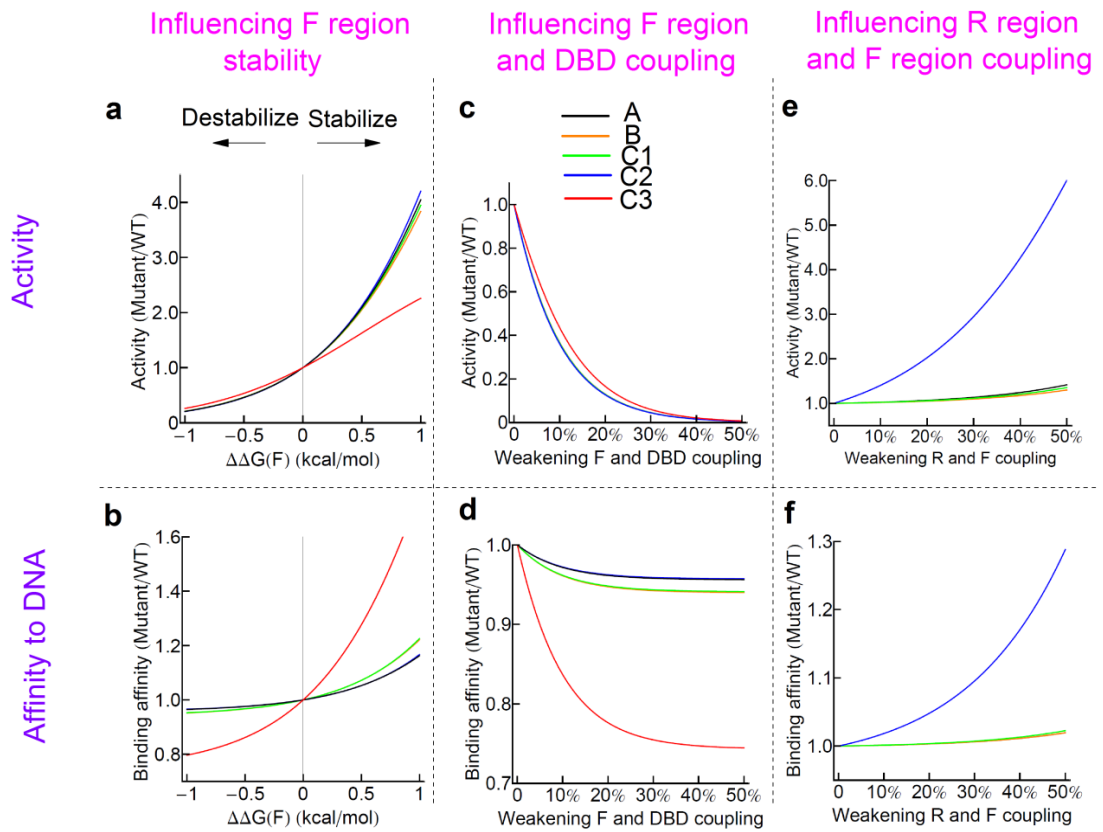


Figure 33 | Predicted influence on transcriptional activity and binding affinity to GRE when mutations are introduced to F domain in different isoforms.

Three possible thermodynamic scenarios may happen when a mutation is introduced into the F domain: influencing F stability (shown in a, b), influencing F domain and DBD coupling (shown in c, d), and influencing R domain and F domain coupling (shown in e, f).

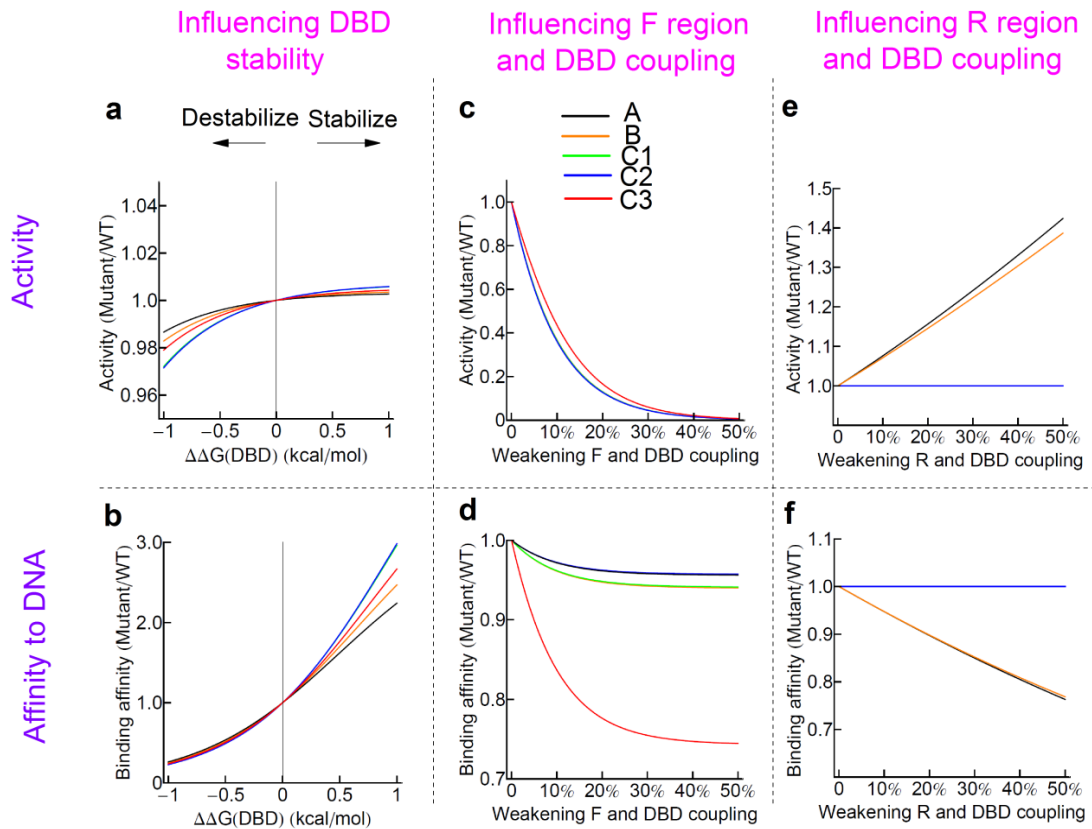


Figure 34 | Predicted influence on transcriptional activity and binding affinity to GRE when mutations are introduced to DBD in different isoforms.

Three possible thermodynamic scenarios may happen when a mutation is introduced into the DBD: influencing DBD stability (shown in a, b), influencing F domain and DBD coupling (shown in c, d), and influencing R and DBD coupling (shown in e, f).

Chapter 9-Molecular Basis for the Unfavorable Coupling between the Regulatory (R) and Functional (F) Domains in GR NTD

9.1 Abstract

In Chapter 2, thermodynamic characterization of GR NTD suggested that the long NTD is composed of two functionally distinct regions, the regulatory (R) and functional (F) domains, which are negatively coupled to each other. To unravel the molecular basis for the allostery between the R and F domains, mutations were introduced into the R domain of different GR translational isoforms and the functional effects were measured by the dual luciferase reporter assay. Single point mutations within segment 90-97, which corresponds to the R domain of the C2 isoform, significantly increase the transcriptional activity of the C2 isoform. This suggested that the segment 90-97 may be involved in the coupling network between the R and F domains. Guided by EAM predictions, the mutational effects of E92Q and E94Q were also tested on A, B and C1 isoforms, which further supported the hypothesis that E92 and E94 are involved in the coupling between the R and F domains. Several mutations in a predicted coiled-coil (119-132) in the F domain showed different effects on the transcriptional activity of the A and C2 compared with the C3 isoform. Remarkably, K120Q and S125A mutations significantly increased transcriptional activity on A and C2 isoforms, however, they did not significantly influence the activity of the C3 isoform, indicating their involvement in mediating the coupling between the F and R domains.

9.2 Introduction

To investigate the allosteric coupling mechanism, in addition to studying the thermodynamics of the coupling, it is also desirable to elucidate the molecular basis of the

coupling. Such information can pave a potential way for the design of allosteric drugs, which may have advantages over orthosteric drugs [63-65]. Previous literature has usually used point mutations and the functional and thermodynamic effects of these mutations were interpreted in the framework of a coupling model [63]. As our experimental system is a transcription factor and the dual luciferase reporter assay has already been developed as our functional assay, in this chapter, we tried to probe the molecular basis of allostery between the R and F domains through this functional assay. Mutational effects are interpreted in the context of the GR EAM, which makes it straightforward and possible to understand the results.

9.3 Materials and methods

Site directed mutagenesis

All site directed mutagenesis work was done with polymerase chain reactions (PCR) with Phusion polymerase (NEB). Complimentary primers were designed to have 10 ~15 nucleotides on each side of the manipulation site, and the calculated T_m value of the primer pair should be within the range of 70 °C to 80°C for the PCR to be effective. All the primers were synthesized by IDT. The PCR was carried out according to the manufacturer's protocol, and including 1.5% DMSO in the reaction is usually helpful. After PCR, 1uL of DpnI (NEB) was added to each 10uL reaction product followed by a 4 hour 37°C incubation to digest the template. After DpnI treatment, the PCR product was transformed to chemically competent, NEB5α cells (made in lab, transformation efficiency higher than 10⁶). Transformed cells were plated on LB agar plates with selective antibiotics and the next morning single colonies were picked up. Plasmids were prepared with a mini prep kit and sequenced to check the correctness.

Western blot

Western blot protocol is the same as in Chapter 5.

Transcriptional activity measurement at the saturated GR vector level

U-2 OS cells were plated on 96 well plates at a density of 3×10^4 cells/well. After 18~24 hours, when the cells reach about 80-90% confluence, X-tremeGENE HP DNA transfection reagent (Roche) mediated chemical transfection was carried out. For the transcriptional activity at saturated GR vector levels, 40ng GRE2-Gluc, 40ng pCluc-miniTK2 and 5ng GR expression vector were co-transfected into U-2 OS cells on a 96-well plate. After 48 hours, *Gaussia* Luciferase activity and *Cypridina* Luciferase activity were measured as described in Chapter 5.

9.4 Results and discussion

Mutations within segment 90-97 dramatically increase transcriptional activity on C2 isoform

As shown in Fig.1, the C2 isoform (GR 90-525) and C3 isoform (GR 98-525) differ by only eight amino acids, MGETETKV, on the C terminal end of the R domain. However, the transcriptional activity of the C2 isoform is more than 5 fold lower than that of the C3 isoform (shown in Fig. 21). As the R and F domains are negatively coupled to each other (discussed in Chapter 6), we hypothesize that this segment, GR 90-97 (MGETETKV) may play an important role in regulating the negative coupling between the R and F domains. Inspection of the GR 90-97 sequence, we note it is highly polar and charged. Single point mutations, E92Q, E92A, T93V, E94Q, E94A, T95V and K96Q, were introduced to the C2 construct. All mutants, except the K96Q, had similar expression levels compared to wild type C2 as shown by the Western blot analysis (Fig.35 upper panel). For all the mutants showed comparable expression levels to the WT GR, transcriptional activities were measured (shown in Fig.35 lower panel). Importantly, all mutants significantly increased the transcriptional activity compared to wild type C2.

Introduction of a mutation to the R domain can produce different effects. They can change the R domain stability, influencing the coupling between R and F domains, and/or affecting the coupling between the R domain and DBD. As discussed in chapter 8, the EAM of GR could be used to make predictions about transcriptional activity and binding affinity changes from each of these scenarios. As discussed in chapter 8, and also shown in the left panel of Fig.36, mutations within the R domain are predicted to have distinguishable effects on different isoforms under each scenario.

In order to distinguish the different scenarios, two of the mutations, E92Q and E94Q, which showed dramatic transcriptional activity increases on the C2 isoform, were also made in the context of the A, B and C1 isoforms. Comparing the transcriptional activity to their wild type, E92Q and E94Q only caused a significant transcriptional activity increase on the C2 isoform (as shown in Fig.36 right panel). This fits the scenario that these two mutations weaken the coupling between the R domain and F domain (as shown in Fig.36 left). Thus, E92 and E94 within the R domain are involved in the R domain and F domain coupling network.

Another possibility is that GR 90-97, may contain the co-repressor binding sites, and mutations may influence co-repressor binding and increase the transcriptional activity compared to wild type C2. To check this possibility, constructs with varied lengths (4aa, 8aa, and 16aa) of flexible linkers preceding the N terminus of the C3 isoform were made and transcriptional activities were tested. As shown in Fig.37, all these constructs showed significantly decreased transcriptional activity compared to the C3 isoform (about 2.5 fold on average), even though they had higher activity than the C2 isoform (about 2 fold increase on average). This suggests that the N terminus of the C3 isoform may have important functional sites, and coupled folding and

binding occurs when it binds to co-regulators. Any additional sequence preceding its N terminus may unfavorably influence its ability to adopt the correct conformation upon folding. However, the naturally existing C2 isoform, with its unique eight amino acids flanking C3, has even lower activity compared to the constructs with flexible linkers on C3. This indicates that there may be some specific interaction between these eight amino acids and the F domain.

Of course these results cannot preclude the possibility that the eight amino acids, MGETETKV, contain co-repressor binding sites. The fact that no ligand or cofactor has yet to be identified as binding to this region, and the EAM presented here precisely describes the impact of mutated residues involved in the coupling, suggest the model is at least plausible, if not probable.

Mutations on the F domain have different functional effects on A and C2 isoforms compared with C3 isoform

To explore the molecular basis of allostery between the R and F domains in more detail, we sought to determine which residues in the F domain are involved in coupling to the R domain. As the F domain is greater than 300 amino acids in length, it is not possible to exhaustively explore mutational space in this domain. Fortunately, in the F domain, secondary structure predictions can pinpoint local element that may form structure. One such region (GR 119-132) is predicted to be a coiled-coil. Coiled-coil is a structure motif where 2 to 7 helices interact with each other through hydrophobic and electrostatic interactions to form a coil [66]. As discussed above, the influence of flexible linkers on the C3 isoform suggests that there may be some important functional regions near the N terminus of the F domain, and these regions may be also coupled to the eight amino acids unique to the C2 isoform. As coiled coil regions are often used for interdomain or intermolecular interaction, they represent for the residues that may mediate coupling.

Selected single point mutations (L119A, K120Q, L122A, S125A and L129A), which were predicted to reside on the surface of the coiled coil, were carried out on the C2 and A isoforms to check whether any are involved in coupling to the R domain. All mutations were also carried out on the C3 isoform to gauge the influence on the stability of the F domain. As predicted by the EAM (left panel of Fig.38), if a mutation only destabilizes the F domain or weakens the coupling between the F domain and DBD, it would cause transcriptional activity to decrease for all three isoforms, with substantially greater decreases for the A and C2 isoforms compared to the C3 isoform. If a mutation were to weaken the coupling between the R and F domains, it will cause a large transcriptional activity increase for the C2 isoform, less of a transcriptional activity increase for the A isoform, and no transcriptional activity change for the C3 isoform.

Among the panel of mutations (Fig.38 right panel), L119A, L122A and L129A are predicted to significantly destabilize the coiled-coil region, resulting a transcriptional activity decrease for all the isoforms, with the decrease being minimal for the C2 isoform. K120Q and S125A mutations are predicted to slightly increase the coiled-coil propensity, and consistently, these mutations did not significantly change the transcriptional activity on the C3 isoform. However, on the C2 isoform, these two mutations caused a significant transcriptional activity increase. The results on all these five mutations are inconsistent with the scenario that these mutations only influence the F domain stability or F domain and DBD coupling. Instead, these results suggest that the predicted coiled coil region on the N terminal end of the F domain may be involved in the allosteric coupling between the R and F domains, with residues K120 and S125 within this region contribute significantly to the observed negative coupling.

9.5 Conclusions

By testing selected mutational effects on different GR translational isoforms, and comparing the results to predictions of the EAM for GR, we were able to probe the molecular basis for the negative coupling between the R domain and F domain in the NTD. Our results suggest that some residues within segment 90-97, are involved in the coupling between the R and F domains. Similarly, some residues (in particular K120 and S125) located within the predicted coiled-coil (segment 119-132) in the F domain contribute significantly to coupling the F domain to the R domain.

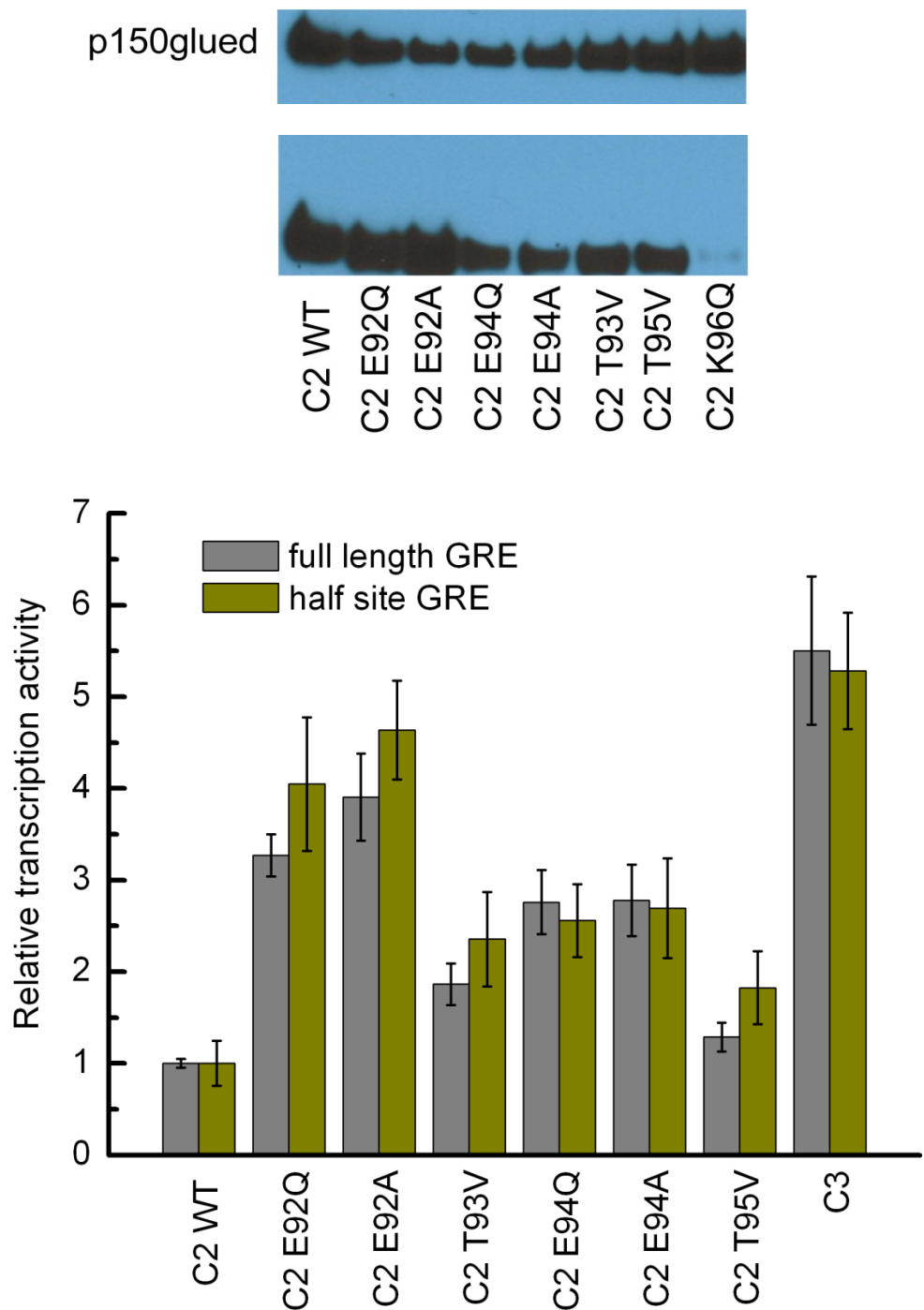


Figure 35 | Most point mutations within segment 90-97 significantly increase the transcriptional activity compared to wild type C2.

The upper panel is a Western blot testing the expression level of all the mutants made within segment 90-97. For the mutants with reasonable expression levels, transcriptional activity was tested through dual luciferase reporter assays as shown in the lower panel.

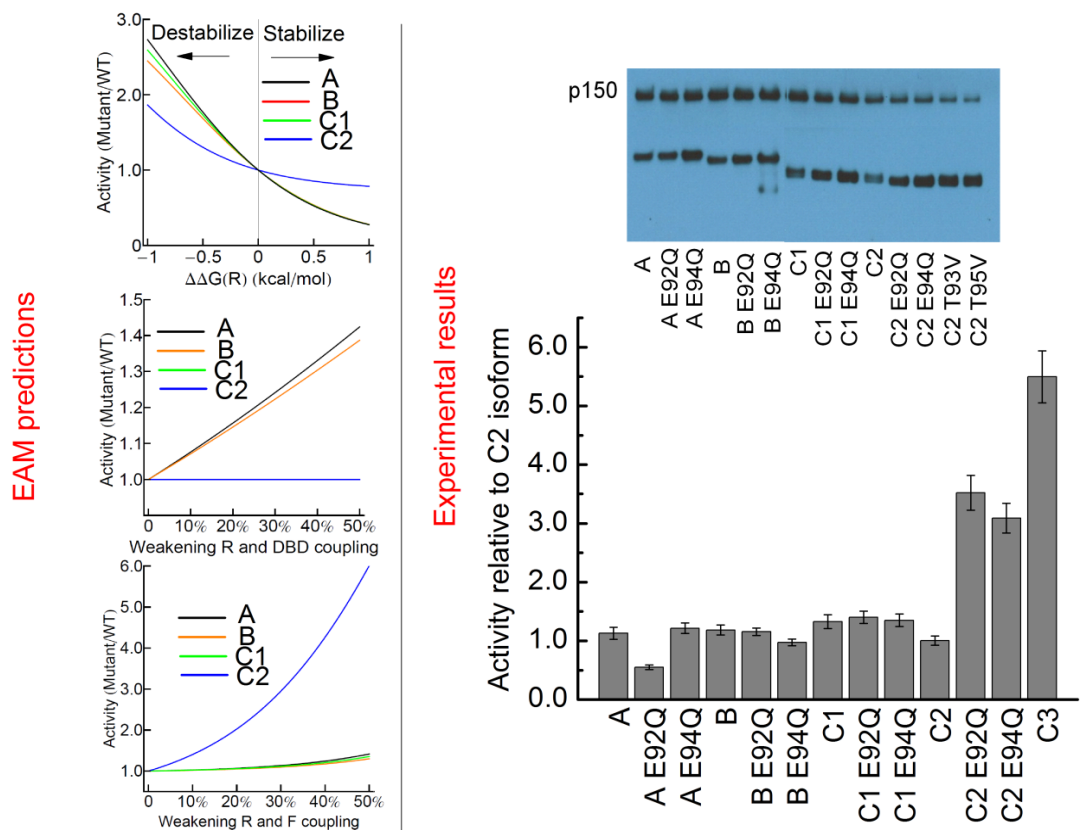


Figure 36 | Mutations within R domain on A, B, C1 and C2 isoforms to identify residues involved in the R and F coupling network.

The effect on transcriptional activity when mutations are introduced to the R domain on A, B, C1, C2 and C3 isoforms can be predicted based the EAM model for three scenarios (I: influencing R domain stability; II: decreasing R and DBD coupling; III: decreasing R and F coupling) as shown in the left panel. The bar graph shows the influence of E92Q and E94Q mutations on the activity of A, B, C1, and C2 isoforms measured by a luciferase reporter assay. The transcriptional activity of each construct is expressed relative to the wild type C2 isoform.

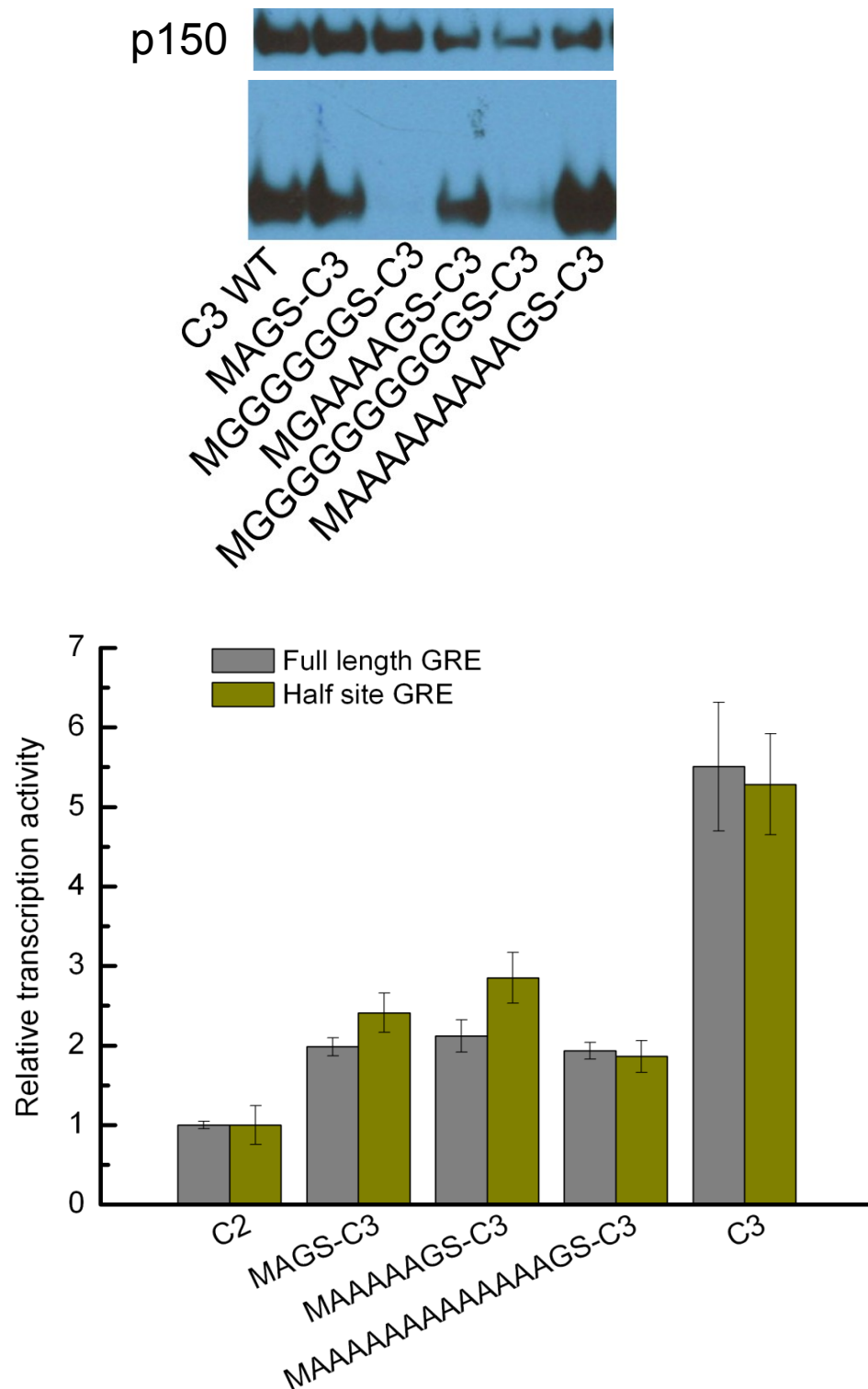


Figure37 | Flexible linkers with varied length flanking the N termini of C3 significantly decrease the transcriptional activity.

The upper panel is a Western blot testing the expression level of some constructs with flexible linkers preceding the N terminus of C3. For the constructs with a reasonable expression level, transcriptional activity was tested through the dual luciferase reporter assay shown in the lower panel.

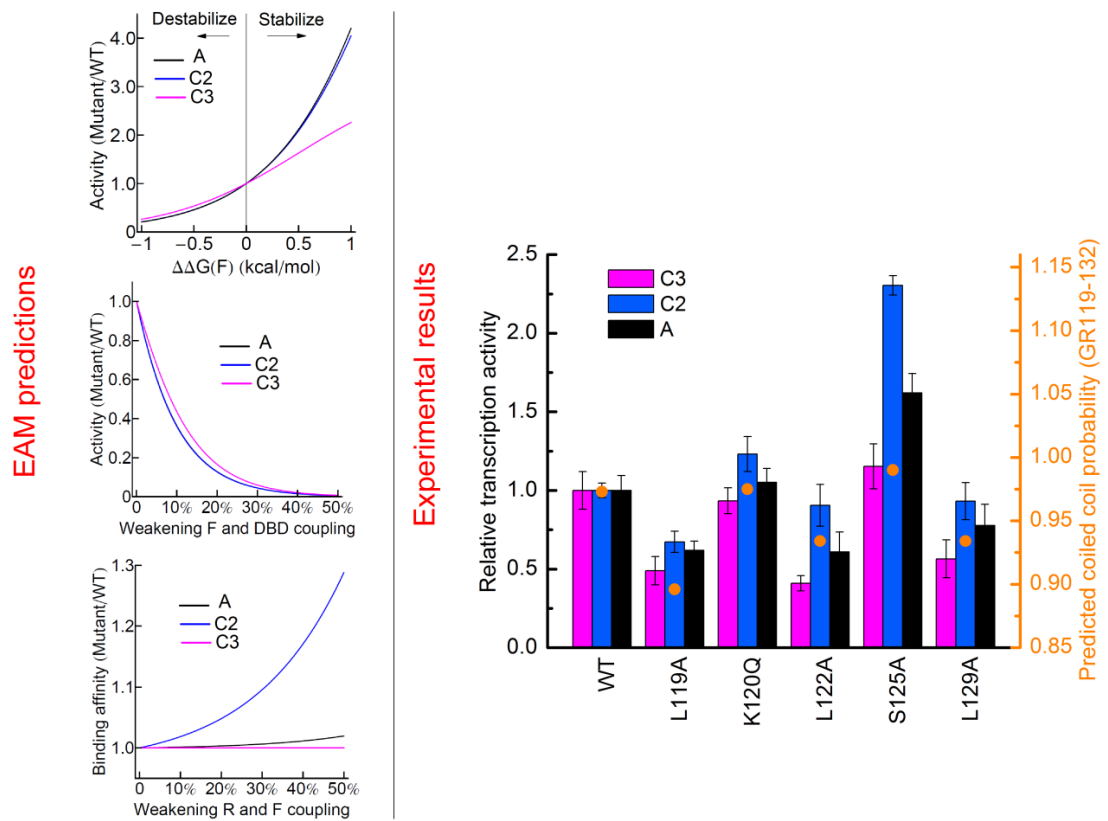


Figure 38 | Mutations within F domain on A, C2 and C3 isoforms to identify residues involved in the R and F coupling network.

The effect on transcriptional activity when mutations are introduced to the F domain on A, C2 and C3 isoform can be predicted based on the EAM model for three scenarios (influencing F domain stability, weakening F and DBD coupling or weakening R and F coupling) as shown in the left panel. Single point mutants (L119A, K120Q, L122A, S125A and L129A) in the functional domain where a coiled-coil structure is predicted, either show indistinguishable activity compared with wild type (K120Q and S125A) or significantly lower activity for the C3 isoform (L119A, L122A and L129A). It is consistent with the coiled-coil probability prediction (right dot corresponding to the right axes). However, on the C2 isoform, K120Q and S125A show a significant activity increase compared to wild type. And L119A, L122A and L129A mutants showed a less significant transcriptional activity decrease compared to C3. Based on the predicted scenarios, these mutations not only influence the coiled-coil propensity, but also contribute to the coupling between the R and F domains.

Chapter 10-Molecular Basis of the Favorable Coupling between the Functional (F) domain in the NTD and DBD

10.1 Abstract

To explore the molecular basis of allosteric coupling between the F domain and the DBD, single point mutations and chimeric constructs were introduced to the C3 isoform and tested using the luciferase reporter assay. The results support the hypothesis that there is favorable coupling between the F domain and DBD. In the context of the EAM, interpreting the functional effects of single point mutations indicated that some residues in the N terminal domain of the D2 isoform (GR 331-420) are responsible for coupling the F domain to the DBD. Furthermore, on the DBD, involved in coupling to F domain is a surface composed of residues after the first recognition helix.

10.2 Introduction

It is well-established that DNA response elements often act as allosteric regulators of transcription factors. This is also the case for GR, in which binding the GR response element (GRE) to DBD favors recruitment of co-regulators to the activation function 1 (AF1) site in the NTD [18]. To investigate the allosteric coupling between the NTD and DBD in GR, in chapter 3, we carried out the TMAO induced folding experiments on the NTD of the A and C3 isoforms in the presence and in absence of DBD. It showed that the DBD stabilizes the F domain in the NTD, which indicates there is favorable coupling between the F domain and DBD.

To explore the molecular basis of the coupling, functional assays were carried out on constructs designed to interrupt the coupling. Transcriptional activity measurements of a chimeric construct confirmed the favorable coupling between the F domain and DBD. In addition,

interpreting the functional effect of single point mutations in the context of the EAM allowed us to demonstrate the molecular basis of the allostery.

10.3 Materials and methods

Constructs

Plasmid PJ603-C3 NTD-Gal4 DBD is to express the chimeric construct linking the F domain (NTD of C3 isoform) and the DBD (1-147) of the yeast transcription factor Gal4 (Uniprot, P04386) with a 10 amino acid flexible linker (GTGGSGGSGS). It was codon optimized and synthesized by DNA 2.0.

Plasmid Gal4 UAS -Gluc, to express secreted *Gaussia* Luciferase under the control of two tandem Gal4 upstream activating sequences (Gal4 UAS), was made by inserting an oligonucleotide containing two tandem Gal4 UAS, 5'- aattCGGAGTACTGTCCTCCG aggaattcagCGGAGTACTGTCCTCCG-3', into the EcoRI and BamHI sites of the pGluc-miniTK vector (NEB).

All the single point mutations on GR constructs were made by site directed mutagenesis [60] as described in chapter 9.

Western blot

Western blot protocol is the same as described in Chapter 5

Transcriptional activity measurement

Protocol is the same as described in Chapter 5.

10.4 Results and discussion

Chimeric construct linking F domain to yeast transcription factor Gal4's DBD showed significantly decreased transcriptional activity compared to natural GR construct

As discussed in chapter 8 and shown in Fig.33, mutagenesis within the F domain of different isoforms will not distinguish between an influence on F domain stability and a perturbation of F domain and DBD coupling; the two scenarios have similar effects on both transcriptional activity and binding affinity to GRE. Thus to pinpoint the residues within the F domain that are allosterically coupled to the DBD a control construct was designed, in which the F domain is tethered to the DNA binding domain of Gal4, a yeast transcription factor. Transcriptional activity of the chimeric construct was measured on the Gal4 upstream activating sequence driving luciferase expression. As shown in Fig.39, the two-domain construct of natural GR C3 isoform has about 30-fold higher transcriptional potency than the chimeric construct. This functional analysis suggests that GR DBD stabilizes the F domain and favors the binding of co-regulators, supporting the notion of a favorable coupling between F domain of the NTD and DBD.

Mutagenesis on F domain reveals some residues involved in the coupling network

Multiple single point mutations were introduced into the F domain of the C3 isoform in an attempt to identify residues involved in coupling. Several mutations (D335N, S370A, D372N, D373N and T376V), showed significantly decreased transcriptional activity compared to wild type (as shown in Fig. 40). To check whether the mutations influence F domain stability or co-regulator binding, they were also carried out in the context of a chimeric construct composed of the F domain and the DNA binding domain of the yeast transcription factor Gal4. As shown in Fig.41 right panel, the mutations did not significantly change the transcriptional activity on the chimeric control construct. These results support the notion that multiple residues in the N terminal domain of the D2 isoform (GR 331-420) are responsible for coupling the F domain to DBD, of which

D335, S370, D372, D373 and T376 are a subset.

Mutagenesis on DBD pinpoints some residues involved in coupling to F domain

To identify residues in the DBD involved in coupling to the F domain, sequence conservation analysis was carried out [67] on the DBD of GRs from different species and on the whole steroid hormone receptor (SHR) family (Fig.42). Residues that are more conserved in GRs from different species and less conserved within the whole SHR family were selected for mutations, except when those residues were also in contact with DNA base pairs. As shown in Fig.43, multiple mutations (E450Q, R460T, D462N, I464T and R479G) significantly decreased transcriptional activity of the C3 isoform. As predicted by our GR EAM model (shown in the left panel of Fig.44), mutations influencing DBD stability will significantly decrease the binding affinity to GRE, while not significantly influencing the transcriptional activity. In contrast, mutations influencing the F domain and DBD coupling will significantly decrease the transcriptional activity, while not greatly influencing the binding affinity.

To further determine whether the decreased transcriptional activity of these mutants is due to significantly decreased binding affinity to GRE, dosage curve measurements were carried out for the mutants in the dual luciferase assay as shown in the upper right panel of Fig.44. From these dosage curves, both the maximum transcriptional activity and the binding affinity (represented by EC50) for each mutant can be fitted (as shown in the right bottom panel of Fig.44). Mutations of two residues R460 and D462, which are within the dimerization box of DBD, significantly decreased binding affinity to GRE, with maximum transcriptional activity being comparable to the wild type. Mutations of E450, I464 and R479 did not significantly influence binding to GRE, however, significantly lowered transcriptional activity compared to WT, suggesting these residues

may be involved in the allosteric coupling network between F domain and DBD. Interestingly, all these residues can be mapped on a contiguous surface after the first recognition helix on the DBD. In fact, E450 is the first residue that is defined as the lever arm of GR's DBD, which has been proposed to allosterically transmit signal from DBD to other regulatory surfaces [22, 68]. Here, we found that a potential surface after the first recognition helix on DBD mediates coupling to the F domain in the NTD via residues E450, I464 and R479, which were identified through mutagenesis.

10.5 Conclusions

The molecular basis of the allosteric coupling between F and DBD was further investigated using the functional assays. Single point mutations within the F domain, D335N, S370A, D372N, D373N and T376V, significantly decreased transcriptional activity of the C3 isoform, while maintaining similar activity on the chimeric construct where the F domain and DBD coupling is knocked out. This result suggested that these residues within the F domain may contribute to coupling the F domain to DBD. In addition, E450Q, I464T and R478G mutations within the DBD on the C3 isoform significantly decrease transcriptional activity compared to wild type, and did not change the EC50 significantly. This indicates that these residues are involved in coupling to the F domain.

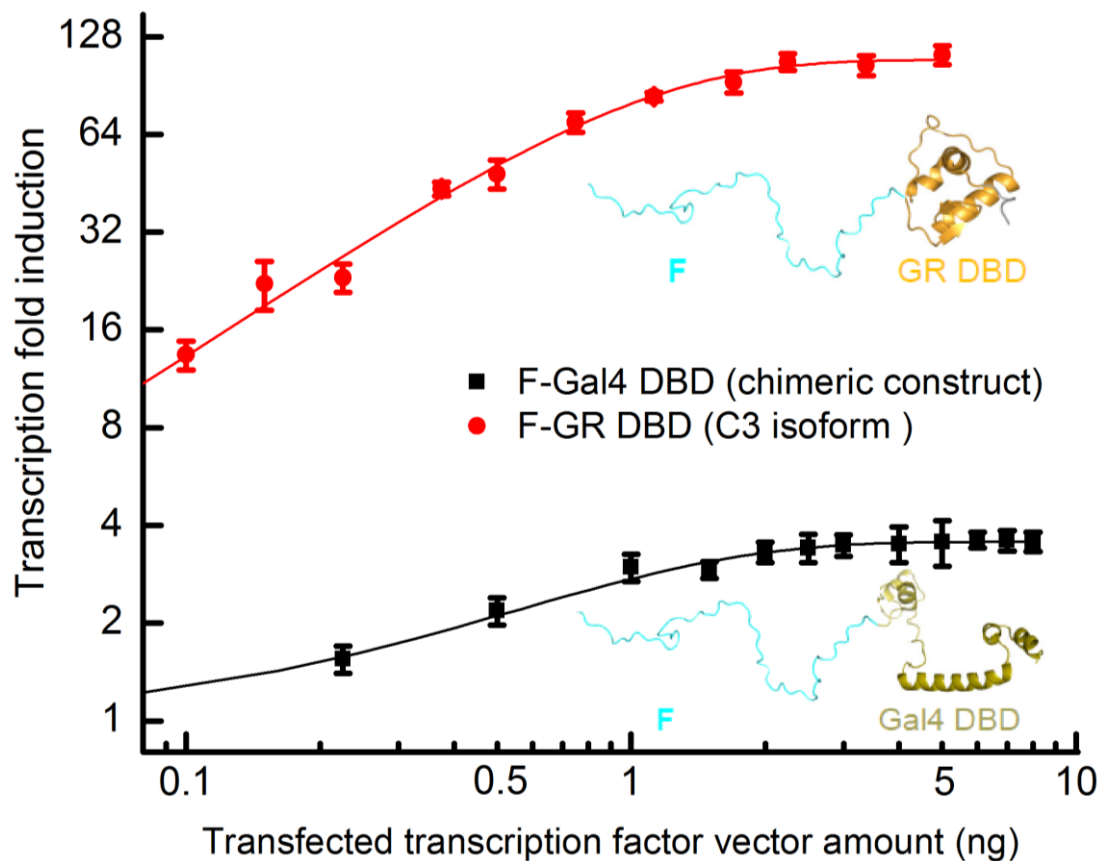


Figure 39 | Domain swapping of GR DBD to Gal4 DBD significantly decrease transcriptional activity.

The luciferase assay dosage curve for C3 isoform (composed of F domain and GR DBD) versus chimeric construct (tethering F domain to yeast transcription factor Gal4's DBD). Changing GR DBD to Gal4 DBD decreased transcriptional activity about 30 fold, suggesting favorable coupling between GR's DBD and F domain.

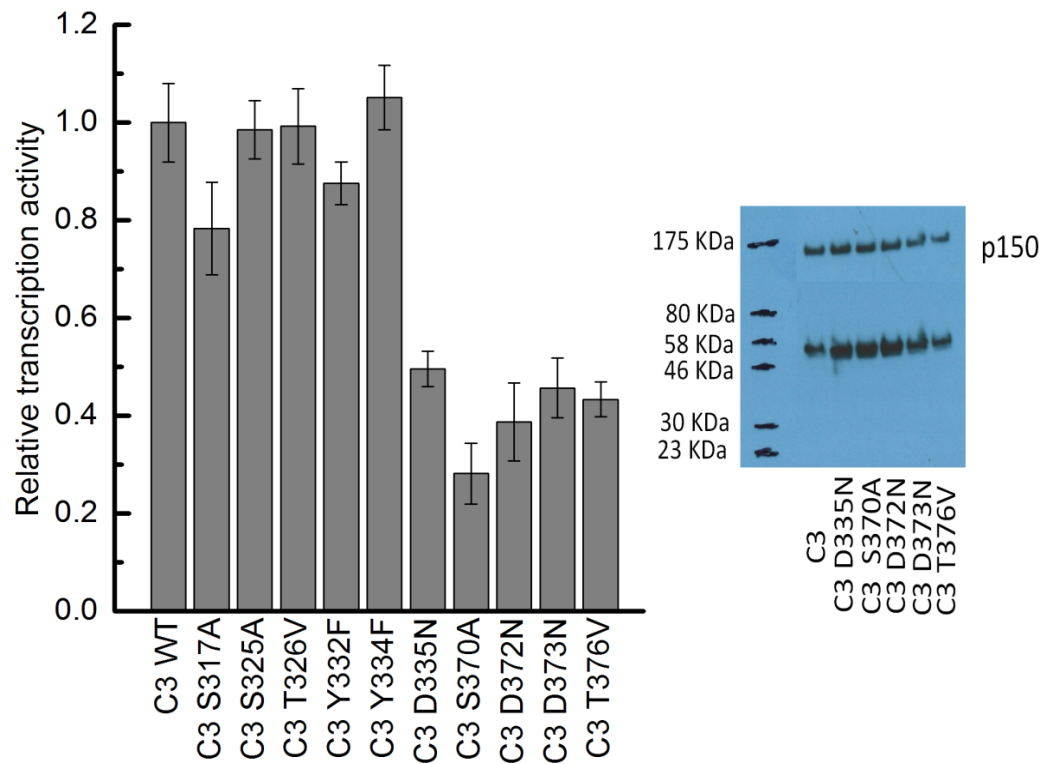


Figure 40 | Transcriptional activity influence of single point mutations within F domain tested on C3 isoform.

For the mutations that significantly decrease the transcriptional activity (D335N, S370A, D372N, D373N and T376V), a Western blot was done to check the protein expression level change as shown in the right.

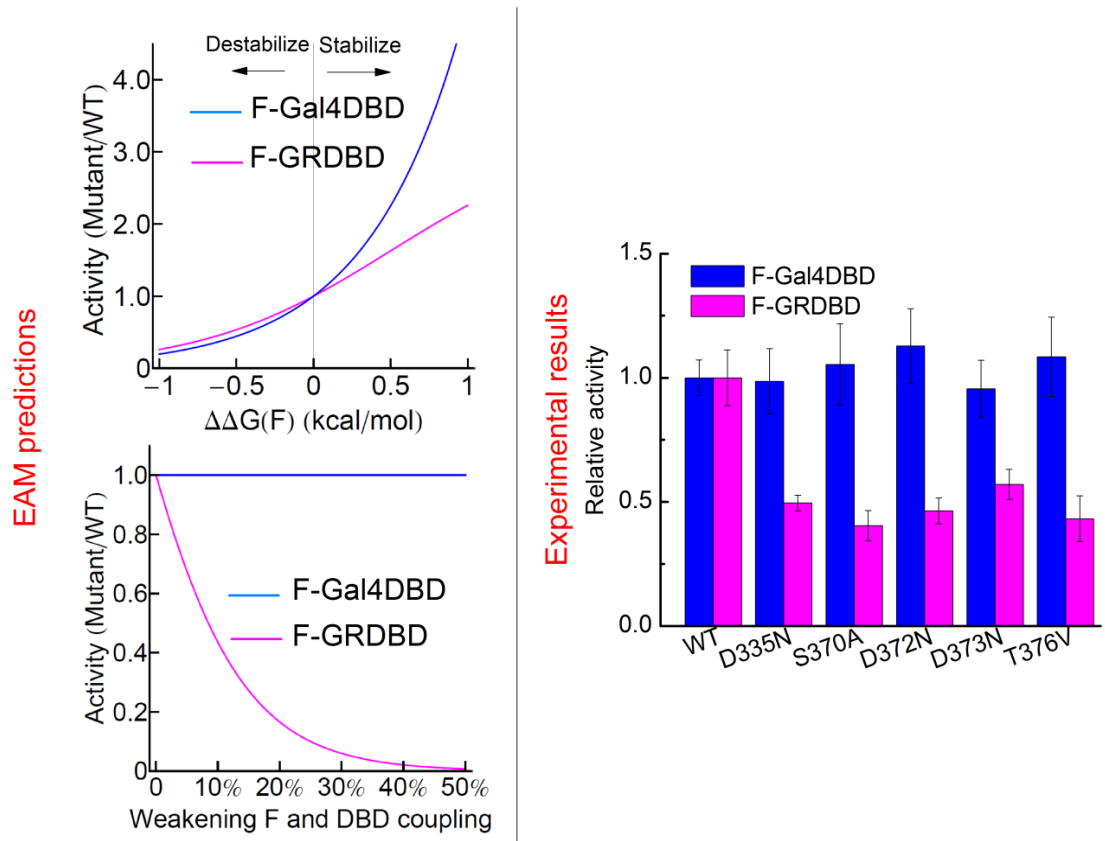


Figure 41 | Mutations within F domain on C3 isoform and chimeric construct (F domain tethered to Gal4 DBD) to identify residues involved in the F and DBD coupling network

The effect on transcriptional activity when mutations are introduced to the F domain on F-GR DBD (C3 isoform) and F-Gal4 DBD (chimeric construct) can be predicted using the EAM model for two scenarios (influencing F domain stability or weakening F and DBD coupling) as shown in the left panel. Relative luciferase activity of the single point mutations in D2 NTD of the functional (F) domain (D335, S370, D372, D373 and T376) tested on C3 NTDDBD (magenta bars) and C3 NTD-Gal4 DBD chimeric construct (blue bars) are shown in the bar graph in the right panel. Interpreting from the predicted scenarios, these residues are involved in the F domain and DBD coupling.

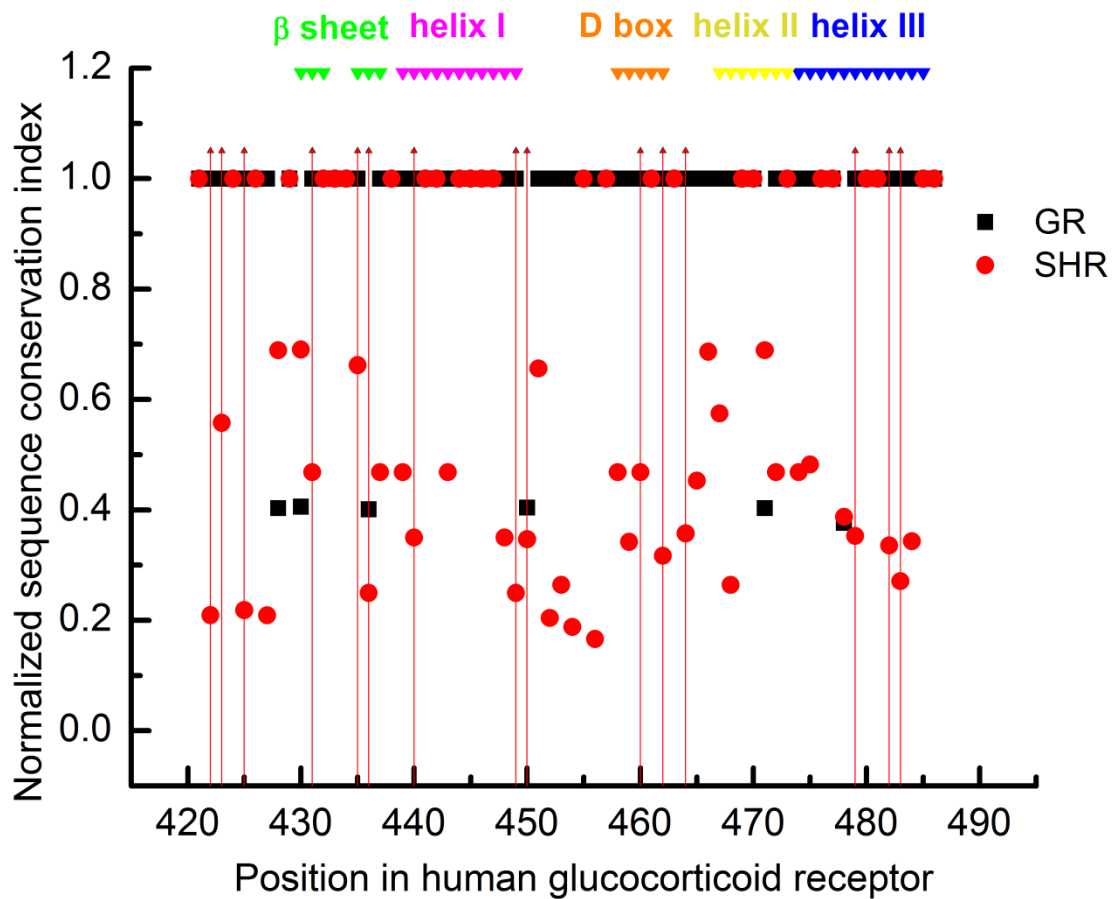


Figure 42 | Sequence conservation analysis of DBD within GRs versus SHRs.

Normalized sequence conservation index versus residue positions based on full-length human GR. The conservation index was calculated from the DBD of GRs in different species and from DBD's of the entire steroid hormone receptor family. Secondary structure of DBD is labeled on top of the graph. The red lines indicated the mutations within DBD carried out in this study.

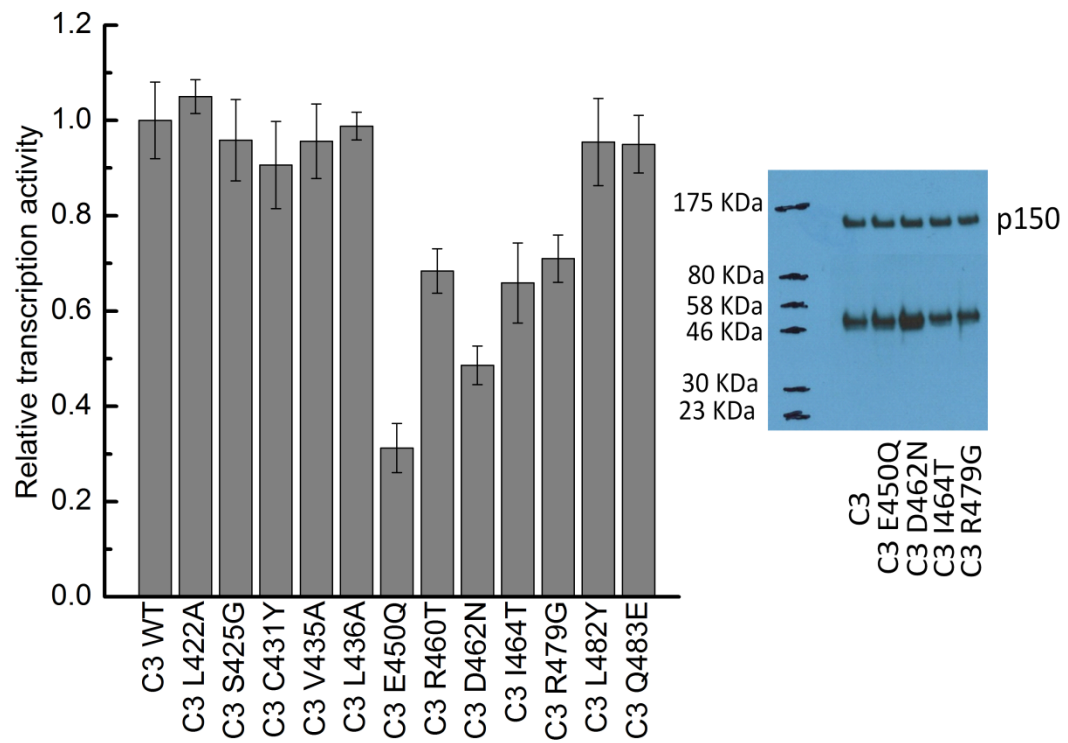


Figure 43 | Influence on transcriptional activity of single point mutations carried out within DBD on C3 isoform.

For the mutations that significantly decrease the transcriptional activity (E450Q, R460T, D462Q, I464T, R479G), a Western blot was carried out and shown in the right panel.

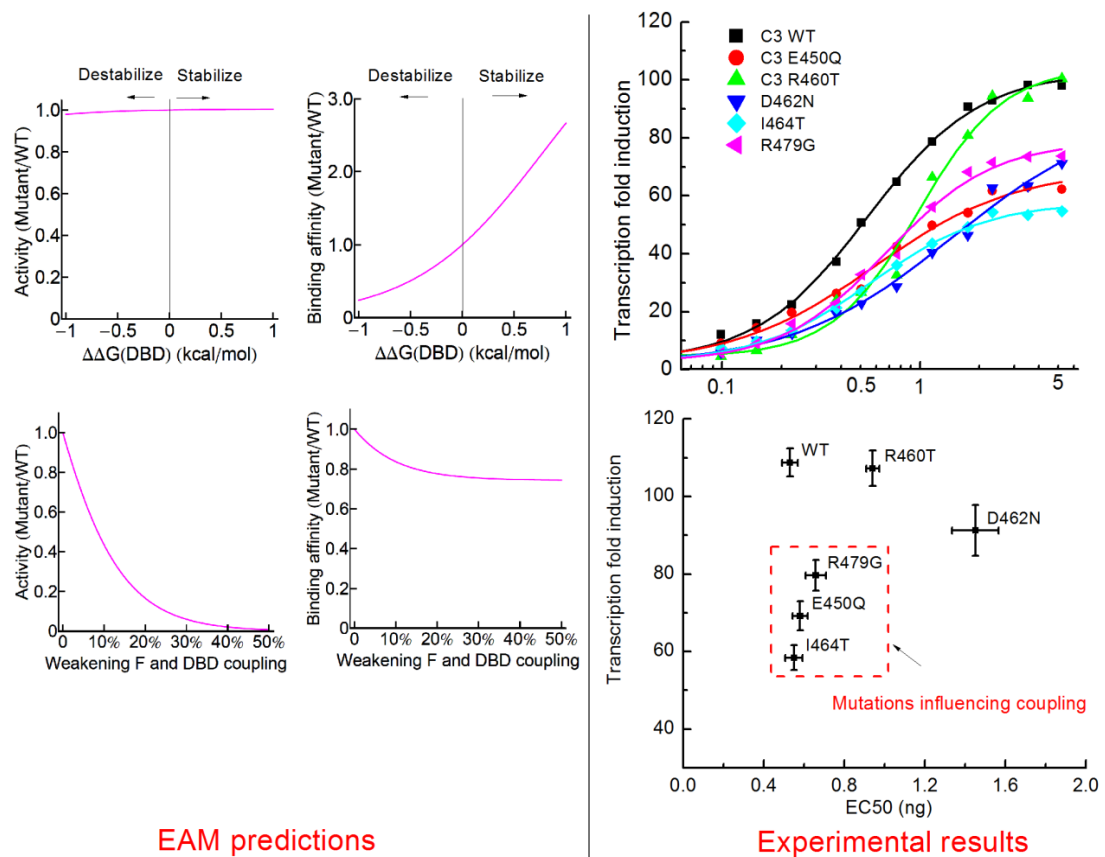


Figure 44 | Mutations within DBD on C3 isoform to identify residues involved in the F and DBD coupling network.

The effect on transcriptional activity and binding affinity when mutations are introduced to the C3 isoform can be predicted based the EAM model for two scenarios (influencing DBD stability or weakening F and DBD coupling) as shown in the left panel. The right top panel shows the luciferase assay dosage curve measured for the mutants (E450Q, R460T, D462N, I464T, and R479G) that showed a significant transcriptional activity decrease compared to WT. From these dosage curves, maximum transcriptional activity and EC50 can be fitted for each construct as plotted in the right bottom panel. Among these, E450Q, I464T and R479G mutations showed significantly decreased transcriptional activity and relatively minor changes in EC50 compared to wild type. Interpreting from the EAM predictions, these residues are involved in the F and DBD coupling network.

Chapter 11- Molecular Basis for the Favorable Coupling between the Regulatory (R) Domain in the NTD and DBD

11.1 Abstract

Interpreting the transcriptional activity and binding affinity data of the eight translational isoforms suggests that there is favorable coupling between the regulatory (R) domain in the NTD and DBD. In chapter 6, by directly linking the R domains of different translational isoforms to the DBD through an eleven amino acid flexible linker, it was found that the R domain unique to A and B isoforms is favorably coupled to the DBD. In this chapter, to probe the molecular basis of R domain and DBD coupling, different segments of the R domain were linked to the DBD to further explore which part of the R domain is involved in the R domain and DBD coupling network. Segment 1-85 was confirmed to be involved in coupling to DBD. To probe which residues on the DBD are coupled to the R domain, mutational effects on binding affinity were compared between R domain linked DBD construct and linker DBD control construct. Results suggest that residues C431, V435 and L436 which reside on a surface before the first recognition helix are involved in coupling the DBD to the R domain.

11.2 Introduction

The R domain in the NTD was found to be favorably coupled to the DBD. To identify the molecular basis of coupling, mutagenesis within the R domain on the A and B isoforms was originally considered. However, based on EAM predictions shown in Fig.32c&d, weakening R domain and DBD coupling will only moderately influence both the transcriptional activity and binding affinity of A and B isoforms. The F domain also exists in the A and B isoforms, making it more complicated to interpret the mutagenesis effect. In this context, a conjoined construct that is

more straightforward to study R domain and DBD allostery was designed. As discussed in chapter 6, this conjoined construct, in which R domain was directly linked to DBD through a flexible linker, showed significantly increased binding affinity to GRE compared with the DBD construct. This confirmed the favorable coupling between the R domain and DBD and verified that the R domain and DBD conjoined construct is able to explore the R domain and DBD coupling mechanism. So in this chapter, truncations and mutagenesis were introduced to the conjoined construct or DBD to pinpoint the molecular basis of coupling between R domain and DBD.

11.3 Materials and methods

Constructs

Plasmid PJ603-1-97-11aa-DBD construct to express GR 1-97 linked to DBD with an eleven amino acids GTGGSGGSGGS flexible linker was made by ligating together the PCR product deleting the codons for GR 98-420 from PJ603-A NTDDBD with BamHI and KpnI sticky ends and an oligonucleotide coding for GTGGSGGSGGS. Plasmid PJ603-27-97-11aa-DBD, -86-97-11aa-DBD, -90-97-11aa-DBD, -1-85-11aa-DBD, -1-27-11aa-DBD and -11aa-DBD were made by inserting the codons for each construct amplified from PJ603-1-97-11aa-DBD into the NheI and XhoI site of PJ603 vector.

All the single point mutations on GR constructs were made by site directed mutagenesis [60] as described in chapter 9.

Competitive transfection assay

The same as described in Chapter 5.

11.4 Results and discussion

Segment within R domain potentially involved in the R domain and DBD coupling network

As discussed in chapter 6 shown in Fig. 26, R domains of A and B isoforms are favorably coupled to the DBD. In theory, it is desirable to check which segment or residues within the R domain are involved in the coupling in detail. However, both the stability of the R domain and the coupling energy between R domain and DBD will significantly influence the binding affinity of DBD to GRE (as predicted by EAM model shown in Fig.45 left panel), thus it is difficult to identify which segment is specifically involved in the coupling through mutagenesis. To double check whether the R domain unique to A and B is involved in coupling to DBD, segment 1-85 (which is unique to the R domain of the A and B isoform) and segment 1-27 (which is the difference between the A and B isoforms) were also directly linked to the DBD through the flexible linker. As shown in Fig.45 right panel, linking segment 1-85 to DBD has a similar stabilizing effect on the DBD compared to segment 1-97, which is consistent with linking 86-97 and 90-97 to DBD not having a significant effect. Linking segment 1-27 to DBD showed a stabilizing effect on DBD, which is consistent with segment 1-97 having a stronger stabilizing effect on DBD than segment 1-27. Thus all the current data suggest that some residues within the R domain unique to the A and B isoforms (segment 1-85) may be involved in the coupling to DBD.

Residues within DBD potentially involved in the R domain and DBD coupling network

To identify residues on DBD coupled to the R domain, single point mutations were selected (L422A, S425G, C431Y, V435A, L436A, L482Y and Q483E) that did not cause significant changes in transcriptional activity of the C3 isoform (as shown in Fig.43). All these mutations were introduced to the R_A-linker-DBD construct and linker-DBD construct. As predicted from the EAM model (shown in Fig.46 middle panel), mutations influencing DBD stability will

significantly change the binding affinity of both the two constructs. In contrast, mutations influencing the R domain and DBD coupling will only change the binding affinity of R_A-linker-DBD construct. Competitive transfection assays were carried out for all these constructs as shown in Fig.46 bottom panel. The EC50 for each competitor construct can be fitted from these assays as shown in Fig.47. Mutations (C431Y, V435A and L436A) showed significantly different influences on linker-DBD construct and R_A-linker-DBD construct. These mutations did not significantly influence the binding affinity of linker-DBD-construct, and significantly decreased the binding affinity of R_A-linker-DBD construct. Thus, C431, V435 and L436 within DBD are found to be involved in the allosteric coupling network between R domain and DBD.

11.5 Conclusions

By linking different segments of the R domain to the DBD through a flexible linker and measuring the influence on DNA binding affinity, it was found that segment 1-85, which is unique to A and B isoforms, is involved in mediating the allosteric coupling between R domain and DBD. When comparing the mutagenesis effect on the construct with R domain conjoined to DBD to the DBD, we found that some residues (C431, V435 and L436), which reside on a surface before the first recognition helix, are involved in coupling to the R domain.

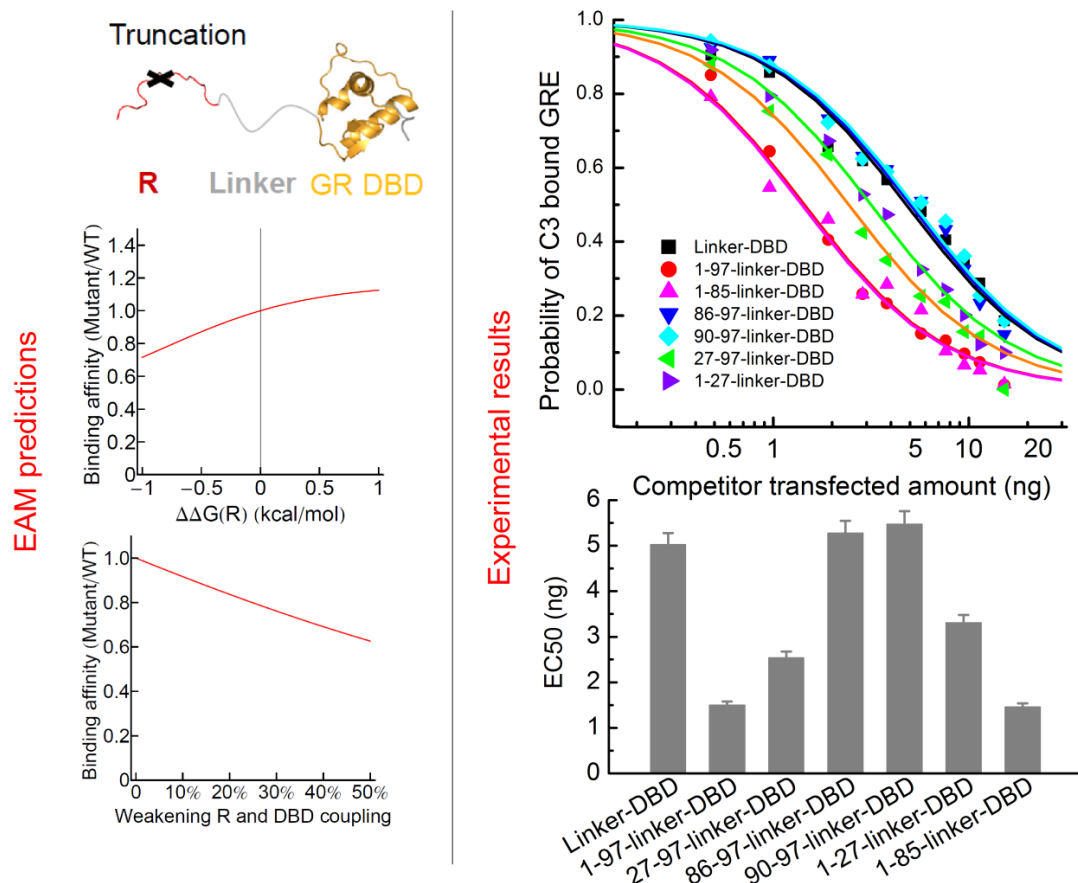


Figure 45 | Truncations of the construct with the R domain joined to the DBD to identify segment involved in the R domain and DBD coupling network.

The effect on binding affinity when changing R domain length can be predicted with an EAM model for two scenarios (influencing R domain stability or weakening R and DBD coupling) as shown in the left panel. Competitive transfection assay curves were measured and shown in the right top panel for: linker-DBD construct, 1-97-linker-DBD construct, 1-85-linker-DBD construct, 86-97-linker-DBD construct, 90-97-linker-DBD construct, 27-97-linker-DBD construct and 1-27-linker-DBD construct. The EC50 can be fitted from these competitive transfection assays for each of these competitor constructs as shown in the right bottom panel. Based on the predicted scenarios, segment 1-85 contributes to the coupling between the R domain and DBD.

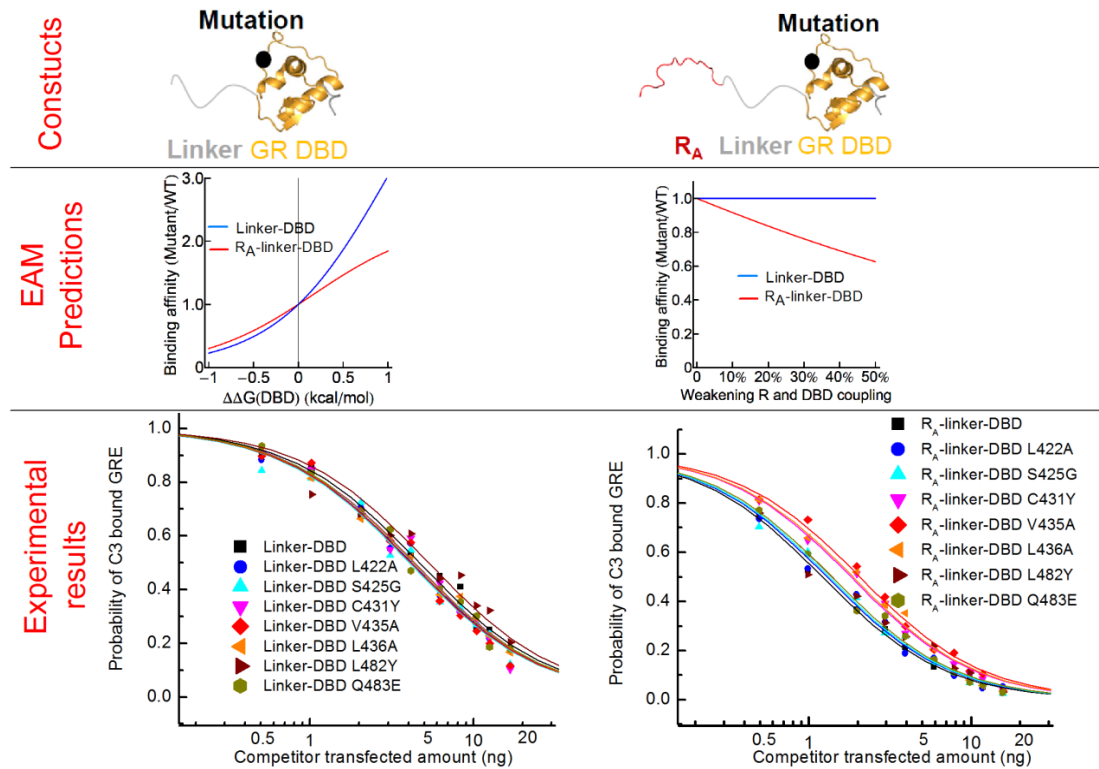


Figure 46 | Mutations within DBD and R_A -DBD construct to identify residues involved in the R and DBD coupling network.

The effect on binding affinity when mutations are introduced to DBD or R_A and DBD conjoined construct can be predicted based on the EAM model for two scenarios (influencing DBD stability or weakening R and DBD coupling) as shown in the left panel. Competitive transfection assays were carried out for single point mutations C431Y, V435A and L436A on both DBD construct and the R_A conjoined to DBD construct as shown in the right middle panel.

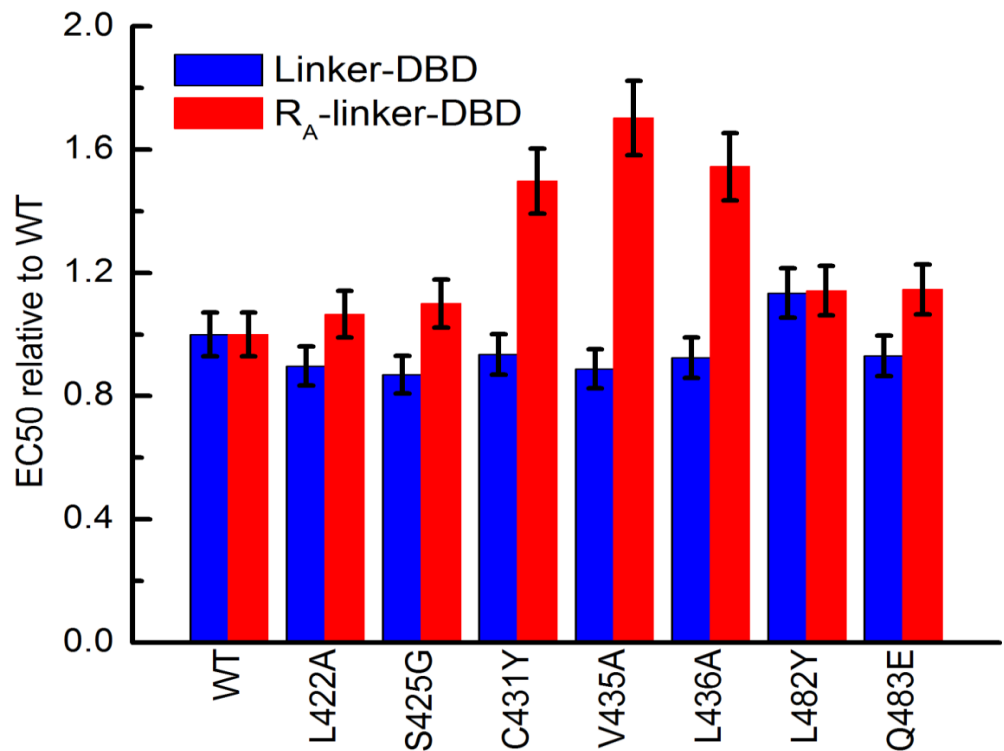


Figure 47 | EC50 of DBD and R_A-DBD constructs' mutants fitted from competitive transfection assay.

EC50 can be fitted from competitive transfection assays for each of these mutants and expressed as a relative value compared to wild type. These three mutations (C431Y, V435A and L436A) showed significantly increased EC50 on the R_A conjoined to DBD construct, and no significant change on the DBD construct. It suggests that they are involved in the coupling network to the R domain.

Chapter 12- Concluding Remarks

12.1 Summary

In this dissertation, allostery mediated by an ID domain in GR was investigated by combining biophysical characterizations, cell based functional assays, and an Ensemble Allosteric Model simulation. The intrinsically disordered NTD was found to be composed of two functionally distinct regions, F domain and R domain, which are unfavorably coupled. Importantly, both the R and F domains are favorably coupled to DBD. The allosteric coupling model was able to provide insight into the reason why there is a lack of correlation between the binding affinity and transcriptional activity of GR translational isoforms. The EAM was invaluable in guiding our investigation into the molecular basis for the allostery (as summarized in Fig.48). Some residues in the NTD of the D2 isoform contribute to coupling the F domain to the DBD, and a surface on the DBD composed of residues after the first recognition helix is involved in coupling to the F domain. For the R domain and DBD allosteric coupling network, segment 1-85 of R domain is engaged in the favorable coupling to DBD. On the DBD, a surface composed of some residues before the first recognition helix is favorably coupled to the R domain. For the negative coupling between the R and F domains, some residues within segment 90-97, which is the R domain of the C2 isoform, and some residues within segment 119-132, which is a predicted coiled-coil in the F domain, are involved in the coupling network.

12.2 Toward future applications

This study suggests a framework in the methodology to study allostery, especially in IDPs. The combination of *in vitro* biophysical studies, cell based functional assays and EAM makes it possible to study a complicated allosteric system composed of long ID regions, as well as

structured domains. Predictions from the EAM guided our mutagenesis efforts by assisting in the selection of the mutation types and isoforms. These prediction-guided experiments were essential in our efforts to interpret the mutational data. The EAM quantitatively captures the variable transcriptional activity and binding affinity of the constitutively active constructs of the eight GR translational isoforms. For example, compared to the full-length A isoform, the higher transcriptional activity and lower binding affinity to GRE of the C3 isoform can be understood. In the literature, the heightened transcriptional activity of the C3 isoform is explained by removal of the steric hindrance effect of the R domain on the functional domain [69]. It is one possible scenario that could contribute to the negative coupling between the R domain and F domain. However, this interpretation does not explain the effect on the DNA binding affinity

The molecular basis of allostery paves a potential way for designing allosteric drugs, which have multiple advantages over orthosteric drugs as reviewed in the literature recently [63-65]. In the SHR family, the NTD almost has no sequence conservation between different members. In this sense, the NTD is a good target for drug development. However, the ID characteristics make it non trivial to target. Elucidating the residues involved in the allosteric coupling network will facilitate drug design.

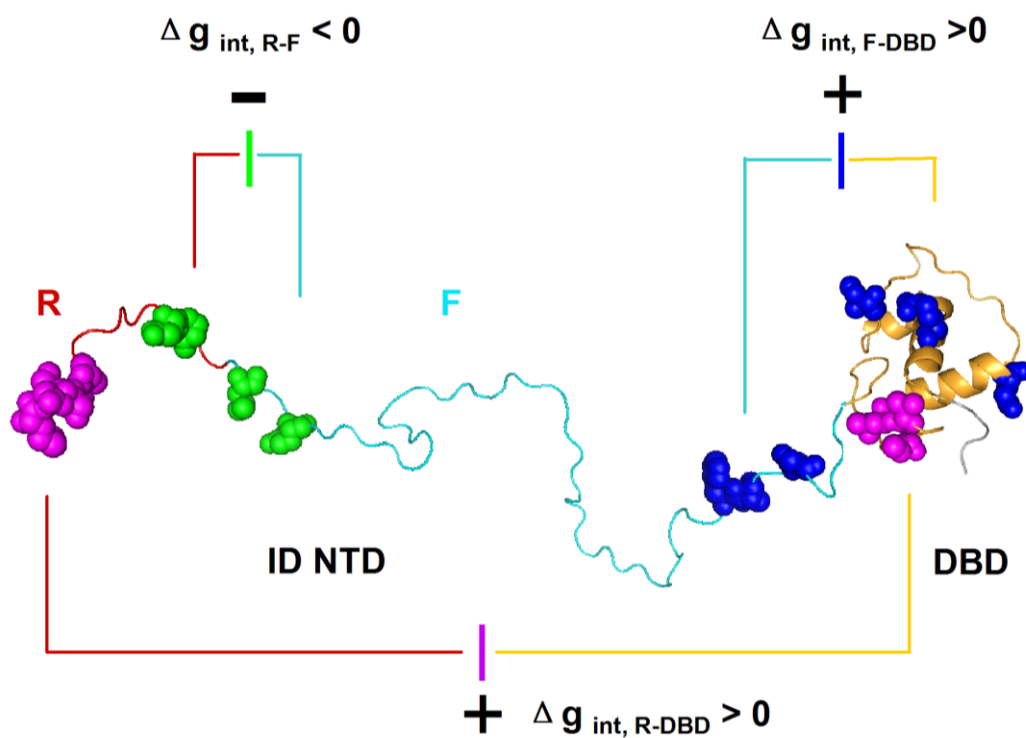


Figure 48 | Cartoon summary of the molecular basis of allosteric coupling in GR constitutively active two-domain constructs

Cartoon summary of the coupling scheme in GR showing the residues in the DBD, R- and F-domains responsible for the inter-domain coupling between each domain.

Reference

1. Changeux, J.P. and S.J. Edelstein, *Allosteric mechanisms of signal transduction*. Science, 2005. **308**(5727): p. 1424-8.
2. Monod, J., J. Wyman, and J.P. Changeux, *On the Nature of Allosteric Transitions: A Plausible Model*. J Mol Biol, 1965. **12**: p. 88-118.
3. Koshland, D.E., Jr., G. Nemethy, and D. Filmer, *Comparison of experimental binding data and theoretical models in proteins containing subunits*. Biochemistry, 1966. **5**(1): p. 365-85.
4. Cui, Q. and M. Karplus, *Allostery and cooperativity revisited*. Protein Sci, 2008. **17**(8): p. 1295-307.
5. Liu, J., et al., *Intrinsic disorder in transcription factors*. Biochemistry, 2006. **45**(22): p. 6873-88.
6. Hilser, V.J. and E.B. Thompson, *Intrinsic disorder as a mechanism to optimize allosteric coupling in proteins*. Proc Natl Acad Sci U S A, 2007. **104**(20): p. 8311-5.
7. Motlagh, H.N. and V.J. Hilser, *Agonism/antagonism switching in allosteric ensembles*. Proc Natl Acad Sci U S A, 2012. **109**(11): p. 4134-9.
8. Wrabl, J.O., et al., *The role of protein conformational fluctuations in allostery, function, and evolution*. Biophys Chem, 2011. **159**(1): p. 129-41.
9. Ward, J.J., et al., *Prediction and functional analysis of native disorder in proteins from the three kingdoms of life*. J Mol Biol, 2004. **337**(3): p. 635-45.
10. Uversky, V.U., Oldfield CJ, Dunker AK, *Showing your ID: intrinsic disorder as an Id for recognition, regulation and cell signaling*. Journal of Molecular Recognition, 2005. **18**: p. 343-384.
11. Gao, J. and D. Xu, *Correlation between posttranslational modification and intrinsic disorder in protein*. Pac Symp Biocomput, 2012: p. 94-103.
12. Romero, P.R., et al., *Alternative splicing in concert with protein intrinsic disorder enables increased functional diversity in multicellular organisms*. Proc Natl Acad Sci U S A, 2006. **103**(22): p. 8390-5.
13. Griekspoor, A., et al., *Visualizing the action of steroid hormone receptors in living cells*. Nucl Recept Signal, 2007. **5**: p. e003.
14. Oakley, R.H. and J.A. Cidlowski, *Cellular processing of the glucocorticoid receptor gene and protein: new mechanisms for generating tissue-specific actions of glucocorticoids*. J Biol Chem, 2011. **286**(5): p. 3177-84.
15. Hilser, V.J. and E.B. Thompson, *Structural dynamics, intrinsic disorder, and allostery in nuclear receptors as transcription factors*. J Biol Chem, 2011. **286**(46): p. 39675-82.
16. Geserick, C., H.A. Meyer, and B. Haendler, *The role of DNA response elements as allosteric modulators of steroid receptor function*. Mol Cell Endocrinol, 2005. **236**(1-2): p. 1-7.
17. Gronemeyer, H. and W. Bourguet, *Allosteric effects govern nuclear receptor action: DNA appears as a player*. Sci Signal, 2009. **2**(73): p. pe34.
18. Kumar, R., et al., *Interdomain signaling in a two domain fragment of the human glucocorticoid receptor*. J Biol Chem, 1999. **274**(35): p. 24737-41.
19. Meijsing, S.H., et al., *DNA binding site sequence directs glucocorticoid receptor structure and activity*. Science, 2009. **324**(5925): p. 407-10.
20. Pandit, S., et al., *Allosteric effects of dexamethasone and RU486 on glucocorticoid*

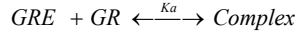
- receptor-DNA interactions*. J Biol Chem, 2002. **277**(2): p. 1538-43.
21. Pfaff, S.J. and R.J. Fletterick, *Hormone binding and co-regulator binding to the glucocorticoid receptor are allosterically coupled*. J Biol Chem, 2010. **285**(20): p. 15256-67.
 22. Watson, L.C., et al., *The glucocorticoid receptor dimer interface allosterically transmits sequence-specific DNA signals*. Nat Struct Mol Biol, 2013. **20**(7): p. 876-83.
 23. Dahlman-Wright, K., et al., *Delineation of a small region within the major transactivation domain of the human glucocorticoid receptor that mediates transactivation of gene expression*. Proc Natl Acad Sci U S A, 1994. **91**(5): p. 1619-23.
 24. Lavery, D.N. and I.J. McEwan, *Structure and function of steroid receptor AF1 transactivation domains: induction of active conformations*. Biochem J, 2005. **391**(Pt 3): p. 449-64.
 25. Lu, N.Z. and J.A. Cidlowski, *Translational regulatory mechanisms generate N-terminal glucocorticoid receptor isoforms with unique transcriptional target genes*. Mol Cell, 2005. **18**(3): p. 331-42.
 26. Li, J., et al., *Thermodynamic dissection of the intrinsically disordered N-terminal domain of human glucocorticoid receptor*. J Biol Chem, 2012. **287**(32): p. 26777-87.
 27. Chen, H., G. Srinivasan, and E.B. Thompson, *Protein-protein interactions are implied in glucocorticoid receptor mutant 465*-mediated cell death*. J Biol Chem, 1997. **272**(41): p. 25873-80.
 28. Bain, D.L., et al., *Glucocorticoid receptor-DNA interactions: binding energetics are the primary determinant of sequence-specific transcriptional activity*. J Mol Biol, 2012. **422**(1): p. 18-32.
 29. Robblee, J.P., M.T. Miura, and D.L. Bain, *Glucocorticoid receptor-promoter interactions: energetic dissection suggests a framework for the specificity of steroid receptor-mediated gene regulation*. Biochemistry, 2012. **51**(22): p. 4463-72.
 30. Tantos, A., K.H. Han, and P. Tompa, *Intrinsic disorder in cell signaling and gene transcription*. Mol Cell Endocrinol.
 31. Tompa, P., *Intrinsically unstructured proteins*. Trends Biochem Sci, 2002. **27**(10): p. 527-33.
 32. Dyson, H.J. and P.E. Wright, *Intrinsically unstructured proteins and their functions*. Nat Rev Mol Cell Biol, 2005. **6**(3): p. 197-208.
 33. Fink, A.L., *Natively unfolded proteins*. Curr Opin Struct Biol, 2005. **15**(1): p. 35-41.
 34. Dyson, H.J. and P.E. Wright, *Coupling of folding and binding for unstructured proteins*. Curr Opin Struct Biol, 2002. **12**(1): p. 54-60.
 35. Babu, M.M., et al., *Intrinsically disordered proteins: regulation and disease*. Curr Opin Struct Biol. **21**(3): p. 432-40.
 36. Kumar, R., et al., *The conformation of the glucocorticoid receptor afl/tau1 domain induced by osmolyte binds co-regulatory proteins*. J Biol Chem, 2001. **276**(21): p. 18146-52.
 37. Baskakov, I. and D.W. Bolen, *Forcing thermodynamically unfolded proteins to fold*. J Biol Chem, 1998. **273**(9): p. 4831-4.
 38. Auton, M. and D.W. Bolen, *Predicting the energetics of osmolyte-induced protein folding/unfolding*. Proc Natl Acad Sci U S A, 2005. **102**(42): p. 15065-8.
 39. McEwan, I.J., et al., *Natural disordered sequences in the amino terminal domain of nuclear receptors: lessons from the androgen and glucocorticoid receptors*. Nucl Recept Signal, 2007. **5**: p. e001.
 40. Hollenberg, S.M. and R.M. Evans, *Multiple and cooperative trans-activation domains of the*

- human glucocorticoid receptor*. Cell, 1988. **55**(5): p. 899-906.
41. Bolen, D.W. and I.V. Baskakov, *The osmophobic effect: natural selection of a thermodynamic force in protein folding*. J Mol Biol, 2001. **310**(5): p. 955-63.
 42. Bolen, D.W., *Protein stabilization by naturally occurring osmolytes*. Methods Mol Biol, 2001. **168**: p. 17-36.
 43. Baskakov, I.V., et al., *Trimethylamine N-oxide-induced cooperative folding of an intrinsically unfolded transcription-activating fragment of human glucocorticoid receptor*. J Biol Chem, 1999. **274**(16): p. 10693-6.
 44. Pace, C.N., et al., *How to measure and predict the molar absorption coefficient of a protein*. Protein Sci, 1995. **4**(11): p. 2411-23.
 45. Tsumoto, K., et al., *Role of arginine in protein refolding, solubilization, and purification*. Biotechnol Prog, 2004. **20**(5): p. 1301-8.
 46. Santoro, M.M. and D.W. Bolen, *Unfolding free energy changes determined by the linear extrapolation method. 1. Unfolding of phenylmethanesulfonyl alpha-chymotrypsin using different denaturants*. Biochemistry, 1988. **27**(21): p. 8063-8.
 47. Wu, P. and D.W. Bolen, *Osmolyte-induced protein folding free energy changes*. Proteins, 2006. **63**(2): p. 290-6.
 48. Wright, P.E. and H.J. Dyson, *Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm*. J Mol Biol, 1999. **293**(2): p. 321-31.
 49. Auton, M. and D.W. Bolen, *Application of the transfer model to understand how naturally occurring osmolytes affect protein stability*. Methods Enzymol, 2007. **428**: p. 397-418.
 50. Ghosh, K. and K.A. Dill, *Computing protein stabilities from their chain lengths*. Proc Natl Acad Sci U S A, 2009. **106**(26): p. 10649-54.
 51. Pradeep, L. and J.B. Udgaonkar, *Osmolytes induce structure in an early intermediate on the folding pathway of barstar*. J Biol Chem, 2004. **279**(39): p. 40303-13.
 52. Henkels, C.H., et al., *Linked folding and anion binding of the Bacillus subtilis ribonuclease P protein*. Biochemistry, 2001. **40**(9): p. 2777-89.
 53. Mello, C.C. and D. Barrick, *Measuring the stability of partly folded proteins using TMAO*. Protein Sci, 2003. **12**(7): p. 1522-9.
 54. Uversky, V.N., *Natively unfolded proteins: a point where biology waits for physics*. Protein science : a publication of the Protein Society, 2002. **11**(4): p. 739-56.
 55. Whitmore, L., et al., *PCDDb: the Protein Circular Dichroism Data Bank, a repository for circular dichroism spectral and metadata*. Nucleic acids research, 2011. **39**(Database issue): p. D480-6.
 56. Han, J.H., et al., *The folding and evolution of multidomain proteins*. Nat Rev Mol Cell Biol, 2007. **8**(4): p. 319-30.
 57. Kassel, O. and P. Herrlich, *Crosstalk between the glucocorticoid receptor and other transcription factors: molecular aspects*. Mol Cell Endocrinol, 2007. **275**(1-2): p. 13-29.
 58. Tannous, B.A., et al., *Codon-optimized Gaussia luciferase cDNA for mammalian gene expression in culture and in vivo*. Mol Ther, 2005. **11**(3): p. 435-43.
 59. Nakajima, Y., et al., *cDNA cloning and characterization of a secreted luciferase from the luminous Japanese ostracod, Cypridina noctiluca*. Biosci Biotechnol Biochem, 2004. **68**(3): p. 565-70.
 60. Liu, J., et al., *A parallel coiled-coil tetramer with offset helices*. Biochemistry, 2006. **45**(51): p.

15224-31.

61. Suel, G.M., et al., *Evolutionarily conserved networks of residues mediate allosteric communication in proteins*. Nat Struct Biol, 2003. **10**(1): p. 59-69.
62. Brown, C.J., A.K. Johnson, and G.W. Daughdrill, *Comparing models of evolution for ordered and disordered proteins*. Mol Biol Evol, 2010. **27**(3): p. 609-21.
63. Maksay, G., *Allostery in pharmacology: thermodynamics, evolution and design*. Prog Biophys Mol Biol, 2011. **106**(3): p. 463-73.
64. Berezhovsky, I.N., *Thermodynamics of allostery paves a way to allosteric drugs*. Biochim Biophys Acta, 2013. **1834**(5): p. 830-5.
65. Nussinov, R. and C.J. Tsai, *Allostery in disease and in drug discovery*. Cell, 2013. **153**(2): p. 293-305.
66. Mason, J.M. and K.M. Arndt, *Coiled coil domains: stability, specificity, and biological implications*. Chembiochem, 2004. **5**(2): p. 170-6.
67. Pei, J. and N.V. Grishin, *AL2CO: calculation of positional conservation in a protein sequence alignment*. Bioinformatics, 2001. **17**(8): p. 700-12.
68. Thomas-Chollier, M., et al., *A naturally occurring insertion of a single amino acid rewires transcriptional regulation by glucocorticoid receptor isoforms*. Proc Natl Acad Sci U S A, 2013. **110**(44): p. 17826-31.
69. Bender, I.K., Y. Cao, and N.Z. Lu, *Determinants of the heightened activity of glucocorticoid receptor translational isoforms*. Mol Endocrinol, 2013. **27**(9): p. 1577-87.

Appendix 1. Derivation of the fitting equation for GRE and GR binding



$$Ka = \frac{C_{boundGRE}}{C_{freeGRE} C_{freeGR}} = \frac{C_{boundGRE}}{(C_{GRE} - C_{boundGRE})(C_{GR} - C_{boundGRE})}$$

$$C_{boundGRE} = \frac{1 + KaC_{GRE} + KaC_{GR} - \sqrt{(1 + KaC_{GRE} + KaC_{GR})^2 - 4Ka^2 C_{GRE} C_{GR}}}{2Ka}$$

$$P_{boundGRE} = \frac{C_{boundGRE}}{C_{GRE}} = \frac{1 + KaC_{GRE} + KaC_{GR} - \sqrt{(1 + KaC_{GRE} + KaC_{GR})^2 - 4Ka^2 C_{GRE} C_{GR}}}{2KaC_{GRE}}$$

$$\begin{aligned} A_{obs} &= A_{freeGRE} \times P_{freeGRE} + A_{boundGRE} \times P_{boundGRE} = A_{freeGRE} \times (1 - P_{boundGRE}) + A_{boundGRE} \times P_{boundGRE} \\ &= (A_{boundGRE} - A_{freeGRE}) \times \frac{1 + KaC_{GRE} + KaC_{GR} - \sqrt{(1 + KaC_{GRE} + KaC_{GR})^2 - 4Ka^2 C_{GRE} C_{GR}}}{2KaC_{GRE}} + A_{freeGRE} \end{aligned}$$

Appendix 2. C code for GR EAM model

/*consider 5% is poised to respond. more detailed binding affinity and activity compare*/

/* How to compile?

gcc -fopenmp allisoform.c -o allisoform -lm */

#include <stdio.h>

#include <string.h>

#include <stdlib.h>

#include <math.h>

#include <time.h>

#include <omp.h>

main()

{

float CRANTDDBD, BindingcompetentANTDDBD;

float CRBNTDDBD, BindingcompetentBNTDDBD;

float CRC1NTDDBD, BindingcompetentC1NTDDBD;

float CRC2NTDDBD, BindingcompetentC2NTDDBD;

float CRC3NTDDBD, BindingcompetentC3NTDDBD;

float BindingcompetentRADBD, BindingcompetentRBDDBD, BindingcompetentDBD;

int

deltaGRE, deltaGf, deltaGd, deltaGrA, deltaGrB, deltaGrC1, deltaGrC2, deltagrAf, deltagrBf, deltagrC1f, deltagrC2f, deltagrAd, deltagrBd, deltagrC1d, deltagrC2d, deltagfd;

int deltaGdhistogram[4]={0,0,0,0};

int deltaGrAhistogram[9]={0,0,0,0,0,0,0,0,0};

int deltaGrBhistogram[9]={0,0,0,0,0,0,0,0,0};

int deltaGrC1histogram[9]={0,0,0,0,0,0,0,0,0};

int deltaGrC2histogram[9]={0,0,0,0,0,0,0,0,0};

int deltagrAfhistogram[6]={0,0,0,0,0,0};

int deltagrBfhistogram[6]={0,0,0,0,0,0};

int deltagrC1fhistogram[6]={0,0,0,0,0,0};

int deltagrC2fhistogram[6]={0,0,0,0,0,0};

int deltagrAdhistogram[10]={0,0,0,0,0,0,0,0,0,0};

int deltagrBdhistogram[10]={0,0,0,0,0,0,0,0,0,0};

int deltagfdhistogram[5]={0,0,0,0,0};

int nthreads, nprocs;

const float R=0.001986; /* Cal/mol -K */

float conval;

int stepsize;

int i;

FILE *outfile;

FILE *outfile2;

/*count=0;*/

```

deltagGRE=-4000;
stepsize = 500; /* Cal/mol */
conval = -1.62349/1000;
deltagrC1d=0;
deltagrC2d=0;
deltaGf=-7000;

/* Determine number of Processors and threads */
#pragma omp single
{
nthreads = omp_get_num_threads();
nprocs = omp_get_num_procs();
printf("Number of Threads per CPU : %5d \n",nthreads);
printf("Number of CPUs : %5d \n",nprocs);
}

outfile=fopen("statisfiedparameter.txt","w");
fprintf(outfile,"deltaGf deltaGd deltaGrA deltaGrB deltaGrC1 deltaGrC2 deltagfd
deltagrC1f deltagrC1f deltagrC1f deltagrC2f deltagrAdd deltagrBd
CRANTDDBD BindingcompetentANTDDBD CRBNTDDBD
BindingcompetentBNTDDBD CRC1NTDDBD BindingcompetentC1NTDDBD
CRC2NTDDBD BindingcompetentC2NTDDBD CRC3NTDDBD
BindingcompetentC3NTDDBD \n");

# pragma omp parallel for default(none) private(i,BindingcompetentRADBD,
BindingcompetentRBDBD,BindingcompetentDBD,CRANTDDBD,
BindingcompetentANTDDBD,CRBNTDDBD, BindingcompetentBNTDDBD,CRC1NTDDBD,
BindingcompetentC1NTDDBD,CRC2NTDDBD,
BindingcompetentC2NTDDBD,CRC3NTDDBD,
BindingcompetentC3NTDDBD,deltaGd,deltaGrA,deltaGrB,deltaGrC1,deltaGrC2,deltagrAf,delta
grBf,deltagrC1f,deltagrC2f,deltagrAd,deltagrBd,deltagfd)
shared(deltaGf,deltaGdhistogram,deltaGrAhistogram,deltaGrBhistogram,deltaGrC1histogram,delt
aGrC2histogram,deltagrAfhistogram,deltagrBfhistogram,deltagrC1fhistogram,deltagrC2fhistogra
m,deltagrAdhistogram,deltagrBdhistogram,deltagfdhistogram,conval,deltagGRE, stepsize, outfile,
outfile2,deltagrC1d,deltagrC2d)

for (deltaGd = -2000 ; deltaGd <= -500 ; deltaGd = deltaGd + 500)
{
for (deltaGrA = -1500 ; deltaGrA <=2500 ; deltaGrA = deltaGrA + 500)
{
for (deltaGrB = -1500 ; deltaGrB <= 2500 ; deltaGrB = deltaGrB + 500)
{
for (deltaGrC1 = -1500; deltaGrC1 <=2500 ; deltaGrC1 = deltaGrC1 + 500)
{

```

```

for (deltaGrC2 = -1500; deltaGrC2 <=2500 ; deltaGrC2 = deltaGrC2 + 500)
{
for (deltagfd = 5000 ; deltagfd <= 7000 ; deltagfd = deltagfd + 500)
{
for (deltagrAf = -3000 ; deltagAf <= -500 ; deltagAf = deltagAf + 500)
{
for (deltagrBf = -3000 ; deltagBf <= -500 ; deltagBf = deltagBf + 500)
{
for (deltagrC1f = -3000 ; deltagC1f <= -500 ; deltagC1f = deltagC1f + 500)
{
for (deltagrC2f = -3000 ; deltagC2f <= -500 ; deltagC2f = deltagC2f + 500)
{
for (deltagrAd = 200; deltagAd <= 2000 ; deltagAd = deltagAd + 200)
{
for (deltagrBd = 200 ; deltagBd <= 2000 ; deltagBd = deltagBd + 200)
{
/*fprintf(stderr,"deltaGd %i\n", deltaGd);*/
CRANTDDBD = exp((conval)*deltagGRE)*( (1+exp((conval)* deltagGRE)) /
(1+exp(( conval)*deltagGRE)+exp((conval)*(deltaGd+deltagfd+deltagrAd))+exp((conval)*(delta
Gd+deltaGf+deltagfd+deltagrAd+deltagrAf))+exp((conval)*(deltaGd+deltagfd+deltaGrA+deltagr
Ad+deltagrAf))+exp((conval)*(deltaGd+deltaGf+deltagfd+deltaGrA+deltagrAd+
deltagrAf))+exp((conval)*(deltaGf+deltagfd+deltagrAf))*( 1+exp((conval)*deltagGRE))+exp((co
nval)*(deltaGrA+deltagrAd+deltagrAf))*(1+exp((conval)*deltagGRE) )+exp((conval)*(deltaGf+
deltagfd+deltaGrA+deltagrAd+ deltagAf))*(1+exp((conval)*deltagGRE))) +
exp((conval)*(deltaGrA+deltagrAd+deltagrAf))*(1+exp(( conval)*deltagGRE)) /
(1+exp((conval)*deltagGRE)+
exp((conval)*(deltaGd+deltagfd+deltagrAd))+exp((conval)*(deltaGd+deltaGf+deltagfd+deltagrA
d+deltagrAf))+exp((conval)*(deltaGd+deltagfd+deltaGrA+deltagrAd+deltagrAf))+
exp((conval)*(deltaGd+deltaGf+deltagfd+deltaGrA+
deltagrAd+deltagrAf))+exp((conval)*(deltaGf+deltagfd+
deltagrAf))*(1+exp((conval)*deltagGRE))+exp((conval)*(deltaGrA+deltagrAd+deltagrAf))*(1+e
xp((conval)*deltagGRE))+exp((conval)*(deltaGf+deltagfd+deltaGrA+deltagrAd+deltagrAf))*(1+
exp((conval)*deltagGRE))));
BindingcompetentANTDDBD = 1 /
(1+exp((conval)*(deltaGd+deltagfd+deltagrAd))+exp((conval)*(deltaGf+deltagfd+deltagrAf))+ex
p((conval)
*(deltaGd+deltaGf+deltagfd+deltagrAd+deltagrAf))+exp((conval)*(deltaGrA+deltagrAd+deltagr
Af))+exp((conval)
*(deltaGd+deltagfd+deltaGrA+deltagrAd+deltagrAf))+exp((conval)*(deltaGf+deltagfd+deltaGrA
+deltagrAd+deltagrAf))+exp((conval)*(deltaGd+deltaGf+deltagfd+deltaGrA+deltagrAd+
deltagrAf))) + exp((conval)*(deltaGf+deltagfd+ deltagAf)) /
(1+exp((conval)*(deltaGd+deltagfd+deltagrAd) )+exp((conval)*(deltaGf+deltagfd+deltagrAf))+e
xp((conval)*(deltaGd+deltaGf+deltagfd+deltagrAd+deltagrAf))+exp((conval)*(deltaGrA+deltagr
Ad+deltagrAf))+exp(( conval)*(deltaGd+deltagfd+deltaGrA+deltagrAd+deltagrAf))+exp((conval)

```


$$\begin{aligned} & \text{elagrC2d+delagrC2f))+exp((conval)*(deltaGd+delagfd+deltaGrC2+delagrC2d+delagrC2f))+} \\ & \text{exp((conval)*(deltaGf+delagfd+deltaGrC2+delagrC2d+delagrC2f))+} \\ & \text{exp((conval)*(deltaGd+deltaGf+delagfd+deltaGrC2+delagrC2d+delagrC2f))) +} \\ & \text{exp((conval)*(deltaGrC2+delagrC2d+delagrC2f))/(1+exp((conval)*(deltaGd+delagfd+} \\ & \text{delagrC2d))+exp((conval)*(deltaGf+delagfd+delagrC2f))+} \\ & \text{exp((conval)*(deltaGd+deltaGf+delagfd+delagrC2d+delagrC2f)) +} \\ & \text{exp((conval)*(deltaGrC2+delagrC2d+delagrC2f))+} \\ & \text{exp((conval)*(deltaGd+delagfd+deltaGrC2+delagrC2d+} \\ & \text{delagrC2f))+exp((conval)*(deltaGf+delagfd+deltaGrC2+} \\ & \text{delagrC2d+delagrC2f))+exp((conval)*(deltaGd+deltaGf+} \\ & \text{delagfd+deltaGrC2+delagrC2d+delagrC2f))) + exp((conval)} \\ & \text{*(deltaGf+delagfd+deltaGrC2+delagrC2d+delagrC2f))/(1+exp((conval)*(deltaGd+delagfd+del} \\ & \text{agrC2d))+exp((conval)*(deltaGf+delagfd+delagrC2f))+exp((conval)*(deltaGd+} \\ & \text{deltaGf+delagfd+delagrC2d+delagrC2f))+exp((conval)*(deltaGrC2+delagrC2d+delagrC2f))+e} \\ & \text{xp((conval)*(deltaGd+} \\ & \text{delagfd+deltaGrC2+delagrC2d+delagrC2f))+exp((conval)*(deltaGf+delagfd+deltaGrC2+delta} \\ & \text{grC2d+delagrC2f))+exp((conval)*(deltaGd+deltaGf+delagfd+deltaGrC2+delagrC2d+delagrC2} \\ & \text{f))}); \end{aligned}$$

$$\text{CRC3NTDDBD} = \frac{\exp((\text{conval}) * \text{delagGRE}) * (1 + \exp((\text{conval}) * \text{delagGRE}))}{(1 + \exp((\text{conval}) * (\text{deltaGd} + \text{delagfd})) + \exp((\text{conval}) * (\text{deltaGd} + \text{deltaGf} + \text{delagfd})) + \exp((\text{conval}) * \text{delagGRE}) + \exp((\text{conval}) * (\text{deltaGf} + \text{delagfd})) * (1 + \exp((\text{conval}) * \text{delagGRE})))};$$

$$\begin{aligned} \text{BindingcompetentC3NTDDBD} &= 1 / (1 + \\ & \exp((\text{conval}) * (\text{deltaGd} + \text{delagfd})) + \exp((\text{conval}) * (\text{deltaGf} + \text{delagfd})) + \exp((\text{conval}) * (\text{deltaGd} + \text{deltaGf} + \text{delagfd}))) \\ &+ \exp((\text{conval}) * (\text{deltaGf} + \text{delagfd})) / (1 + \exp((\text{conval}) * (\text{deltaGd} + \text{delagfd})) \\ &+ \exp((\text{conval}) * (\text{deltaGf} + \text{delagfd})) + \exp((\text{conval}) * (\text{deltaGd} + \text{deltaGf} + \text{delagfd}))); \end{aligned}$$

$$\text{BindingcompetentRADBD} = (1 + \exp(\text{conval} * (\text{deltaGrA} + \text{delagrAd}))) / (1 + \exp(\text{conval} * (\text{deltaGd} + \text{delagrAd})) + \exp(\text{conval} * (\text{deltaGrA} + \text{delagrAd})) + \exp(\text{conval} * (\text{deltaGrA} + \text{deltaGd} + \text{delagrAd})));$$

$$\text{BindingcompetentRBDBD} = (1 + \exp(\text{conval} * (\text{deltaGrB} + \text{delagrBd}))) / (1 + \exp(\text{conval} * (\text{deltaGd} + \text{delagrBd})) + \exp(\text{conval} * (\text{deltaGrB} + \text{delagrBd})) + \exp(\text{conval} * (\text{deltaGrB} + \text{deltaGd} + \text{delagrBd})));$$

$$\text{BindingcompetentDBD} = 1 / (1 + \exp(\text{conval} * \text{deltaGd}));$$

$$\text{/*fprintf(stderr, \"%12.8f \%12.8f \%12.8f \%12.8f \%12.8f \%12.8f \%12.8f \%12.8f\n\", CRBNTDDBD / CRANTDDBD,}$$

$$\begin{aligned} & \text{BindingcompetentBNTDDBD/BindingcompetentANTDDBD, CRC1NTDDBD/CRANTDDBD,} \\ & \text{BindingcompetentC1NTDDBD/BindingcompetentANTDDBD, CRC2NTDDBD/CRANTDDBD,} \\ & \text{BindingcompetentC2NTDDBD/BindingcompetentANTDDBD, CRC3NTDDBD/CRANTDDBD,} \\ & \text{BindingcompetentC3NTDDBD/BindingcompetentANTDDBD);} */ \end{aligned}$$

```

if
( (BindingcompetentRADBD<2*BindingcompetentRBDBD)&&(BindingcompetentRADBD>BindingcompetentRBDBD)&& (BindingcompetentRBDBD>BindingcompetentDBD) &&
(BindingcompetentC3NTDDBD>BindingcompetentDBD) && (BindingcompetentANTDDBD <=0.4)
&&(CRC3NTDDBD/exp((conval)*deltagGRE) >=0.05)&&(CRANTDDBD/CRBNTDDBD < 1)
&& (CRANTDDBD/CRBNTDDBD > 0.68)&& (BindingcompetentANTDDBD >
BindingcompetentBNTDDBD) && (BindingcompetentANTDDBD <
1.5*BindingcompetentBNTDDBD) && (CRANTDDBD/CRC1NTDDBD < 1) &&
(CRANTDDBD/CRC1NTDDBD >0.67)&& (BindingcompetentANTDDBD >
BindingcompetentC1NTDDBD) && (BindingcompetentANTDDBD <
2.5*BindingcompetentC1NTDDBD)&& (CRANTDDBD/CRC2NTDDBD > 1) &&
(CRANTDDBD/CRC2NTDDBD <1.5)&& (BindingcompetentANTDDBD >
BindingcompetentC2NTDDBD) && (BindingcompetentANTDDBD <
2.5*BindingcompetentC2NTDDBD) && (CRANTDDBD/CRC3NTDDBD <0.33) &&
(CRANTDDBD/CRC3NTDDBD >0.2) && (BindingcompetentANTDDBD >
BindingcompetentC3NTDDBD) && (BindingcompetentANTDDBD <
2*BindingcompetentC3NTDDBD)&&(CRBNTDDBD/CRC1NTDDBD > 1) &&
(CRBNTDDBD/CRC1NTDDBD < 1.5)&& (BindingcompetentBNTDDBD >
BindingcompetentC1NTDDBD) && (BindingcompetentBNTDDBD <
2.5*BindingcompetentC1NTDDBD)&& (CRBNTDDBD/CRC2NTDDBD > 1) &&
(CRBNTDDBD/CRC2NTDDBD < 2)&& (BindingcompetentBNTDDBD >
BindingcompetentC2NTDDBD) && (BindingcompetentBNTDDBD <
2.5*BindingcompetentC2NTDDBD) && (CRBNTDDBD/CRC3NTDDBD <0.4) &&
(CRBNTDDBD/CRC3NTDDBD >0.2) && (BindingcompetentBNTDDBD >
BindingcompetentC3NTDDBD) && (BindingcompetentBNTDDBD <
2*BindingcompetentC3NTDDBD) && (CRC1NTDDBD/CRC2NTDDBD > 1) &&
(CRC1NTDDBD/CRC2NTDDBD <2)&& (BindingcompetentC1NTDDBD >
BindingcompetentC2NTDDBD) && (BindingcompetentC1NTDDBD <
2.5*BindingcompetentC2NTDDBD) && (CRC1NTDDBD/CRC3NTDDBD <0.4) &&
(CRC1NTDDBD/CRC3NTDDBD >0.2) && (BindingcompetentC1NTDDBD <
BindingcompetentC3NTDDBD) && (BindingcompetentC1NTDDBD >
0.5*BindingcompetentC3NTDDBD) && (CRC2NTDDBD/CRC3NTDDBD <0.33) &&
(CRC2NTDDBD/CRC3NTDDBD >0.15) && (BindingcompetentC2NTDDBD <
BindingcompetentC3NTDDBD) && (BindingcompetentC2NTDDBD >
0.5*BindingcompetentC3NTDDBD))
{
    fprintf(stderr,"good%i\n",stepsize);
    deltaGdhistogram[(deltaGd+2000)/500]++;
    deltaGrAhistogram[(deltaGrA+1500)/stepsize]++;
    deltaGrBhistogram[(deltaGrB+1500)/stepsize]++;
    deltaGrC1histogram[(deltaGrC1+1500)/stepsize]++;
    deltaGrC2histogram[(deltaGrC2+1500)/stepsize]++;
    deltagfdhistogram[(deltagfd-5000)/stepsize]++;
}

```

```

    deltagrAfhistogram[(deltagrAf+3000)/stepsize]++;
    deltagrBfhistogram[(deltagrBf+3000)/stepsize]++;
    deltagrC1fhistogram[(deltagrC1f+3000)/stepsize]++;
    deltagrC2fhistogram[(deltagrC2f+3000)/stepsize]++;
    deltagrAdhistogram[(deltagrAd-200)/200]++;
    deltagrBdhistogram[(deltagrBd-200)/200]++;
    /*fprintf(stderr,"deltagrBd %i\n", deltagrBd);*/

fprintf(outfile,"%7d %7d %7d %7d %7d %7d %7d %7d %7d %7d %7d %7d %7d %12.8f %12.8f
%12.8f %12.8f %12.8f %12.8f %12.8f %12.8f %12.8f %12.8f %12.8f\n", deltaGf, deltaGd, deltaGrA,
deltaGrB, deltaGrC1, deltaGrC2, deltagfd, deltagrAf, deltagrBf, deltagrC1f, deltagrC2f,
deltagrAd, deltagrBd, CRANTDDBD/exp((conval)*deltagGRE),
BindingcompetentANTDDBD,CRBNTDDBD/CRANTDDBD,
BindingcompetentBNTDDBD/BindingcompetentANTDDBD,CRC1NTDDBD/CRANTDDBD,
BindingcompetentC1NTDDBD/BindingcompetentANTDDBD,CRC2NTDDBD/CRANTDDBD,
BindingcompetentC2NTDDBD/BindingcompetentANTDDBD,CRC3NTDDBD/CRANTDDBD,
BindingcompetentC3NTDDBD/BindingcompetentANTDDBD);
    }

}
}
}
}
}
}
}
}
}
}
}
}

fprintf(stderr,"done with the big loop\n");
fclose(outfile);

outfile2=fopen("histogramreport.txt","w");

for (i=0; i<=3; i++)
{fprintf(outfile2, "deltaG(DBD): ");
fprintf(outfile2, "%5d %100d\n", -2000+i*500, deltaGdhistogram[i]);}

for (i=0; i<=4; i++)
{fprintf(outfile2, "deltagint(F,DBD): ");
fprintf(outfile2, "%5d %100d\n", 5000+i*stepsize, deltagfdhistogram[i]);}

```

```

for (i=0; i<=8; i++)
{fprintf(outfile2, "deltaG(RA): ");
fprintf(outfile2, "%5d %100d\n", -1500+i*stepsize, deltaGrAhistogram[i]);}

for (i=0; i<=8; i++)
{fprintf(outfile2, "deltaG(RB): ");
fprintf(outfile2, "%5d %100d\n", -1500+i*stepsize, deltaGrBhistogram[i]);}

for (i=0; i<=8; i++)
{fprintf(outfile2, "deltaG(RC1): ");
fprintf(outfile2, "%5d %100d\n", -1500+i*stepsize, deltaGrC1histogram[i]);}

for (i=0; i<=8; i++)
{fprintf(outfile2, "deltaG(RC2): ");
fprintf(outfile2, "%5d %100d\n", -1500+i*stepsize, deltaGrC2histogram[i]);}

for (i=0; i<=5; i++)
{fprintf(outfile2, "deltagint(RA,F): ");
fprintf(outfile2, "%5d %100d\n", -3000+i*stepsize, deltagrAhistogram[i]);}

for (i=0; i<=5; i++)
{fprintf(outfile2, "deltagint(RB,F): ");
fprintf(outfile2, "%5d %100d\n", -3000+i*stepsize, deltagrBhistogram[i]);}

for (i=0; i<=5; i++)
{fprintf(outfile2, "deltagint(RC1,F): ");
fprintf(outfile2, "%5d %100d\n", -3000+i*stepsize, deltagrC1histogram[i]);}

for (i=0; i<=5; i++)
{fprintf(outfile2, "deltagint(RC2,F): ");
fprintf(outfile2, "%5d %100d\n", -3000+i*stepsize, deltagrC2histogram[i]);}

for (i=0; i<=9; i++)
{fprintf(outfile2, "deltagint(RA,DBD): ");
fprintf(outfile2, "%5d %100d\n", 200+i*200, deltagrAdhistogram[i]);}

for (i=0; i<=9; i++)
{fprintf(outfile2, "deltagint(RB,DBD): ");
fprintf(outfile2, "%5d %100d\n", 200+i*200, deltagrBdhistogram[i]);}

fclose(outfile2);
}

```

Bibliography

Institution & Location	Dates Attended	Degree (if applicable)	Conferred (mm/yyyy)	Field of Study
Tianjin University, Tianjin, China	09/2000	B.S.	06/2004	Pharmaceutical Engineering
Tianjin University, Tianjin, China	09/2004	M.S.	03/2007	Pharmaceutical Engineering
University of Texas Medical Branch at Galveston, TX	01/2009		05/2010	Molecular Biophysics
Johns Hopkins University, Baltimore, MD	06/2010	Ph.D.	06/2014	Biophysics

Positions/Employment, Memberships and Honors

01/2008–01/2009 Research Associate II, University of Texas Medical Branch,
Department of Pathology, Galveston, TX

Research project: Studying the effect of simultaneous inhibition of ERK and RB pathway on metastatic melanoma.

Peer Reviewed Publications

1. Wu XJ, Li J, Yuan YJ, Liu CX. Progress and perspective in application of bio-informatics to metabolomics and metabonomics. Journal of Chemical Industry and Engineering. 2005; 56(10):1819-1825.
2. Li J, Wu XJ, Liu CX, Yuan YJ. Application of new method for data processing in metabonomic studies. Acta Pharmaceutica Sinica. 2006; 41(1): 47-53.
3. Li J, Xu M, Yang Z, Li A, Dong J. Simultaneous inhibition of MEK and CDK4 leads to potent apoptosis in human melanoma cells. Cancer Investigation. Cancer Invest. 2010; 28

(4):350-6.

4. Schrank TP, Elam WA, **Li J**, Hilser VJ. Strategies for the thermodynamic characterization of linked binding/local folding reactions within the native state application to the LID domain of adenylate kinase from Escherichia coli. *Methods Enzymol.* 2011; 492: 253-82.
5. Motlagh HN *, **Li J***, Thompson EB, Hilser VJ. Interplay between allostery and intrinsic disorder in an ensemble. *Biochem Soc Trans.* 2012; 40(5):975-80. (*: equal contribution)
6. **Li J**, Motlagh HN, Chakuroff C, Thompson EB, Hilser VJ. Thermodynamic dissection of the intrinsically disordered N-terminal domain of human glucocorticoid receptor. *J Biol Chem.* 2012; 287(32):26777-87.
7. Hegde ML, Tsutakawa SE, Hegde PM, Holthauzen LM, **Li J**, Oezguen N, Hilser VJ, Tainer JA, Mitra S. The Disordered C-Terminal Domain of Human DNA Glycosylase NEIL1 Contributes to Its Stability via Intramolecular Interactions. *J Mol Biol.* 2013; 425(13):2359-71
8. Motlagh HN, Wrabl JO, **Li J**, Hilser VJ. The Ensemble Nature of Allostery. *Nature*, 2014 (508): 331-339
9. **Li J**, Wrabl JO, Hilser VJ. Intrinsically Disordered Protein: A Thermodynamic Perspective. Book chapter in *Computational Approaches to Protein Dynamics From Quantum to Coarse-Grained Method*, edited by Dr. Monika Fuxreiter, published by Taylor & Francis. In press.
10. **Li J**, White JT, Saavedra HG, Wrabl JO, Motlagh HN, Liu KX, Sowers J, Trina AS, Thompson EB, Hilser VJ. Parallel Tuning of Activation and Repression in Disorder-Mediated Allostery. (Research article under revision in *Nature*)